DATA MINING METHODS APPLIED TO HEALTHCARE PROBLEMS

A Dissertation
Presented to
The Academic Faculty

by

Sofia Espinoza

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in the
School of Industrial and Systems Engineering

Georgia Institute of Technology
August 2012

DATA MINING METHODS APPLIED TO HEALTHCARE PROBLEMS

Approved by:

Dr. Brani Vidakovic, Advisor
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Paula Edwards, Co-advisor
HIMformatics, LLC

Dr. Melinda Millard-Stafford
School of Applied Physiology
Georgia Institute of Technology

Dr. Yajun Mei
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. David Goldsman
School of Industrial and Systems Engineering
Georgia Institute of Technology

Date Approved: June 26, 2012

To my Mom, Dad, and David,
for giving me the freedom to choose my own path,
for their love and support.

ACKNOWLEDGEMENTS

I wish to thank, first of all, Dr. Brani Vidakovic, my advisor, for his continuous guidance, understanding, encouragement and support, and in particular, for believing in me all the way until the completion of my dissertation work. I would also like to thank Dr. Paula Edwards for graciously accepting to be my co-advisor and having served as a mentor and professional role model during my entire time in the Ph.D. program. I also want to make extensive my gratitude to the rest of my dissertation committee members, Dr. Melinda Millard-Stafford, Dr. Yajun Mei and Dr. David Goldsman for their time in serving on my thesis committee and for providing constructive comments and suggestions that greatly helped to improve this thesis.

Thank you also to my friends, those in Atlanta and outside, who made my life at graduate school more liveable. In particular, thank you to my roommates, Russi and Samer, for making the effort to get to know me and appreciate all my quirks.

Finally, I want to thank my dear family, who are the source of my strength. I would not have gotten where I am without their unconditional support and sacrifice.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

VOCs	Volatile Organic Compounds
CHOA	Children's Healthcare of Atlanta, the pediatric hospital system that served as the setting for the study on medication errors
EG	Egleston, the academic-based hospital at CHOA
SR	Scottish Rite, the community-based hospital at CHOA
GCA	General Care Area
ICU	Intensive Care Unit
EMR	Electronic Medical Record
eMAR	Electronic Medication Administration Record
CPOE	Computerized Provider Order Entry
ADEs	Adverse Drug Events
PONV	Post Operative Nausea and Vomiting experienced by patients after surgery

SUMMARY

Growing adoption of health information technologies is allowing healthcare providers to capture and store enormous amounts of patient data. In order to effectively use this data to improve healthcare outcomes and processes, clinicians need to identify the relevant measures and apply the correct analysis methods for the type of data at hand. In this dissertation, we present various data mining and statistical methods that could be applied to the type of datasets that are found in healthcare research. We discuss the process of identification of appropriate measures and statistical tools, the analysis and validation of mathematical models, and the interpretation of results to improve healthcare quality and safety.

We illustrate the application of statistics and data mining techniques on three real-world healthcare datasets. In the first chapter, we develop a new method to assess hydration status using breath samples. Through analysis of the more than 300 volatile organic compounds contained in human breath, we aim to identify markers of hydration. In the second chapter, we evaluate the impact of the implementation of an electronic medical record system on the rate of inpatient medication errors and adverse drug events. The objective is to understand the impact on patient safety of different information technologies in a specific environment (inpatient pediatrics) and to provide recommendations on how to correctly analyze count data with a large amount of zeros. In the last chapter, we develop a mathematical model to predict the probability of developing post-operative nausea and vomiting based on patient demographics and clinical history, and to identify the group of patients at high-risk.

CHAPTER 1 : INTRODUCTION

The growing adoption of information technologies in healthcare and the availability of more patient data and related healthcare variables provide new opportunities for using analytics to impact health outcomes. Nowadays, most healthcare organizations have some type of healthcare information technology in place (e.g. EMR, CPOE, eMAR), which allows them to measure and capture patient data in real time and at the point the care. As a result, providers have at their disposition enormous amounts of data, which they need to turn into useful information and knowledge that can gear towards improving health outcomes.

Statistical methods such as predictive modeling and data mining are very useful to analyze and interpret the type of datasets found in healthcare research. The research presented here endeavors to expand the knowledge in this area by examining different approaches to the analysis of three healthcare problems. In the first study, we use data mining methods to assess hydration status using breath samples. Through analysis of the more than 300 volatile organic compounds contained in human breath, we aim to identify markers of hydration. For the second problem, we develop a mathematical model to evaluate the impact of an electronic medical record system on the occurrence of medication errors and adverse drug events in a pediatric inpatient setting. In the third study, we apply predictive modeling to estimate the patients' probability of developing post-operative nausea and vomiting based on patient demographics and clinical history. Finally, a more practical objective of this dissertation is to advice practitioners which statistical model works well for a specific context or data.

CHAPTER 2 : ASSESSMENT OF HYDRATION STATUS THROUGH BREATH ANALYSIS

2.1 Literature review

The literature review will explore current methods for assessment of human hydration, advances on breath analysis and from that, describe the research needs in the area.

2.1.1 Hydration assessment

About 63% of the human body mass is water, which is essential for metabolism, temperature regulation and other physiological processes that contribute to good health. Dietary variations, illnesses, strenuous labor, exercise, among other daily activities can alter the normal water requirements for an individual. Even a minimal dehydration of -1% or -2% of body mass can have a negative impact on cognitive function, alertness and exercise performance [1]. At the extreme, acute dehydration, either intentional (e.g., in sports competitions to reach a certain weight) or unintentional, presents a significant health risk that might even result in death [2]. However, increasing evidence indicates that mild dehydration may also account for the development of various morbidities [3]. At the same time, good hydration has been shown to reduce the risk of chronic diseases such as constipation, exercise asthma, hyperglycemia, among others [4]. Thus, the importance of proper hydration and of using reliable and accurate methods to monitor and evaluate even small changes in human hydration levels becomes evident [1, 2].

The following definitions will be used in this chapter [1]:

- Total body water (TBW) is the fluid that occupies intracellular and extracellular spaces.
- Extracellular volume (ECV) is the fluid outside of cells, including the interstitial fluid and plasma water, and comprises about 0.2 L/kg (24.9%) of body mass.
- Intracellular volume (ICV) is the fluid within tissue cells, comprising about 0.4 L/kg (38.4%) of body mass.
- Plasma volume (PV) is the liquid portion of the blood, which makes up about 5% of body mass.
- Interstitial fluid (ISF) is the fluid located in the spaces between tissue cells with a chemical composition similar to that of lymph; it is usually calculated as ECV-PV, and makes up about 21% of body mass.
- An electrolyte is any compound that, in solution, conducts electricity and is decomposed (electrolyzed) by it; i.e., an ionizable substance in solution.
- Osmolality is the concentration of a solution expressed in miliosmoles of solute particles per kilogram of water.
- Euhydration can be defined as “normal body water content” [5]
- Dehydration refers to the process of losing water from the body [6]
- Rehydration is the process of gaining body water [6]
- Hyperhydration refers to the state that exists when ingested fluid temporarily increase total body water above the average basal level [5]
- Hypohydration is a state of being in negative water balance (water deficit) [6]
- Urine specific gravity (U_{sg}) is a measure of the ratio between the density of urine and the density of water [7]

Current researched hydration assessment measures include changes in body mass (BM), blood indices (plasma osmolality (P_{osm}), sodium concentration), urine osmolality (U_{osm}), urine specific gravity (U_{SG}), urine color (U_C) and bioelectrical impedance analysis [1, 6, 8]. Urinary measures have demonstrated being more sensitive than other methods [6], as well as relatively easy to measure [9], but P_{osm} is better at measuring rapid and small changes in hydration status [10]. In search of more practical and noninvasive markers of hydration, the relationship between saliva indices (flow rate, osmolality, spinability) and dehydration has recently been also investigated [11-13]. If no instrument or technical expertise is available, perception of thirst can be taken as an approximation of hydration status [1]. For all these methods, specific threshold values indicating levels of dehydration have been established [2].

Selection of the appropriate hydration assessment method will depend not only on the sensitivity and accuracy needed, but also on the technical requirements and cost of the method [6]. For instance, since high accuracy and precision are essential for laboratory tests, change in body mass would not be a reliable measure of hydration status in that setting [14, 15]. On the other hand, for estimation of hydration status of athletes exercising in the heat, such practical and inexpensive method is deemed accurate and reliable enough [5, 16]. Likewise, markers for hydration status will be different for a static and for a dynamic assessment. In a recent study by Cheuvront et al., only P_{osm} was recommended for static dehydration assessment, while P_{osm}, U_{sg} and BM were all valid in the setting of dynamic hydration assessment [17]. In particular, U_{sg} is the most commonly used to assess hydration status in athletes, as long as muscle mass is taking into account [18]. For any of these methods, it is important to remember that the reported changes in

hydration levels will heavily depend on the initial definition of euhydration [6], and in some cases, even individual euhydrated measures are required [13]. Finally, factors that may affect the response of specific biomarkers include drug or vitamin consumption, food and fluids intake, changes in posture and even how dehydration occurs [18-20].

Despite all the research done in this area, there is still no gold standard for measuring hydration status in all situations and populations [1, 5, 9]. Moreover, the current values that define hydration levels for the various hydration indices are mostly based on expert consensus opinion and not on formal research studies. Only recently, researchers are trying to provide reference values for hydration levels based on studies. For instance, a 12-day study of 59 healthy males by Armstrong et al attempted a statistical approach for providing reference euhydration values and values for extreme cases of hydration [21]. Additionally, hydration biomarkers usually measure dehydration by observing acute changes during exercise in heat and sports [10, 16, 22, 23], but rarely in more "common" condition such as inadequate fluid intake.

2.1.2 Breath analysis and VOCs

Human breath contains nitrogen, oxygen, carbon dioxide, water, inert gases and more than 1000 volatile organic compounds (VOCs). These VOCs can be divided into different chemical classes including hydrocarbons, alcohols, ketones, aldehydes, esters and heterocycles. Unlike NO and other inorganic gases, VOCs are mainly blood-borne, and can reflect a normal physiological biochemical process or a pathological condition in an individual. Recently, due to technical developments in the separation and identification of exhaled substances, the potential of breath analysis as a diagnostic and

monitoring tool in clinical practice and for environmental exposure assessment has greatly increased [24-27].

One of the most interesting applications of breath analysis is the identification of molecular biomarkers of disease [28]. Correlations between concentrations of a single or a set of VOCs and several clinical conditions, such as lung diseases, oxidative stress, and metabolic disorders have been reported in several recent studies [27, 29-31]. The results indicate that biomarkers in the breath can detect gastrointestinal and liver diseases [31], cancers [32], namely lung cancer [33-35], hepatocellular carcinoma [36], pulmonary tuberculosis [37], asthma [38, 39], COPD [39, 40], and type 2 diabetes [41]. In addition, identified biomarkers can also be potentially used for the noninvasive measurement of circulating variables such as plasma glucose (for diabetes management) [42]. Linear and nonlinear multivariate statistical methods have been effective in determining associations between biomarkers and disease occurrences [33, 35, 37, 43, 44].

Exhaled breath contains both, alveolar and “dead space” air. The latter is the ambient air from the upper airway, which dilutes the concentration of the VOCs. Alveolar breath comes directly from the lungs and has been in contact with blood. Since the concentrations of VOCs in the breath are already very low, then it is recommended to collect only the alveolar breath. Another type of contamination in a breath sample comes from exogenous compounds in the ambient air, captured during inspiration. A correction method to minimize these contaminants is to subtract the substance concentrations in the inspired air from the concentrations in the exhaled breath. This difference is called “alveolar gradient”, and its sign indicates whether a particular VOC is endogenous or exogenous [25, 45].

Due to their low concentrations (ppmv to pptv), VOCs have to go through preconcentration and desorption (usually thermal desorption). Separation of the volatile compounds can be done by different methods, being gas chromatography (GC) the most common. Likewise, several detection methods can be used in GC to identify the compounds, and these include: flame ionization detection (FID), mass spectrometry (MS), and ion mobility spectrometry (IMS). Currently, GC/MS is the standard method for determining the composition of VOCs in exhaled breath [25, 26, 45].

Breath tests have some advantages over blood or urine tests because they are noninvasive, easily repeated and present no risk or discomfort for the patient. Furthermore, the composition of breath samples closely reflects the concentration of biological substances in the blood, which would eliminate the need for multiple blood samples [25, 27]. Recent studies even suggest that breath VOC measurements could provide a more consistent measure for investigating underlying physiological function or pathology than single blood measurements [28, 46]. Furthermore, the dynamic behavior of specific VOCs is also being investigated with the hopes of using them for assessments of hemodynamics, pulmonary function and gas exchange patterns [47].

Current limitations in the application of breath analysis in clinical practice include: lack of better understanding of the relationships between breath biomarkers and pathological conditions, lack of standardization and normalization of procedures for the sampling, preparation and analysis of breath samples [25, 27].

2.1.3 Summary of Key Research Needs and specific Aims

From the literature review, we can extract the following main points:

1. Dehydration has negative consequences in our overall health

2. (Accurate) Assessment of hydration status is difficult
3. There is no unique golden standard for measuring hydration status
4. More common methods require collection of bodily fluids
5. Most methods have poor intra- and inter-personal reproducibility
6. Breath tests are noninvasive and easily repeated

In this chapter, we proposed breath analysis as a new method to measure hydration status by analyzing the behavior of VOCs in exhaled breath. We started with an initial exploratory research to assess whether hydration status in adults can be measured through breath analysis. Analyses were conducted from a sample of over 40 healthy subjects, ages 18 to 40. Data mining methods were used to analyze the breath samples and classify hydrated vs. dehydrated individuals.

2.2 Methodology

2.2.1 Data Collection

The collection of breath samples was realized by Dr. Melinda Millard-Stafford research group from the Department of Applied Physiology at Georgia Tech. Human VOCs from the subjects' breath, as well as other hydration measures, were collected for assessment of hydration status. The de-identified dataset was then provided to us for statistical analysis.

Subjects

The subjects were volunteers from the Georgia Institute of Technology campus. Forty-six healthy adult subjects, 23 males and 23 females, ages 18-40 years old were assessed from July 2010 to March 2011. Subjects were excluded from the study for the

following conditions: diabetes, hypothyroid, hyperthyroid, claustrophobia, current weight loss diet, significant weight loss or gain (5% of body weight) in the last three months, liver disease such as hepatitis or cirrhosis, pregnancy, or the inability to refrain from physical labor or exercise for two consecutive days during the study. Informed written consent was obtained prior to initiation of the study as approved by the Georgia Tech Institution Review Board. A within-subject research design was used where each person acted as their own control.

Study Protocol

The study took place over three consecutive days involving a total of five sessions, in which three conditions of hydration status were assessed:

1) Day 1 – Visit 1: AM free living: Prior to the first visit, subjects were told to fast during the 8 hours prior to the visit, although they were allowed to drink water.
2) Day 1 – Visit 2: PM free living: The hydration period started after this visit. The subjects were given instructions for recording their diet and begin the hydration protocol (12ml/kg of body weight during the evening and next day along with normal fluid/food intake).
3) Day 2 – Visit 3: Euhydrated 24-hours following PM free living: After collecting the breath samples, the subjects were given the dehydration protocol. They followed a dry food diet and the only allowable liquids were less than 20ml if needed for medication.
4) Day 3- Visit 4: 16 hours into dehydration.
5) Day 3 – visit 5: 24 hours into dehydration.

During each session, breath samples and a variety of other measures were performed.

Breath samples were collected by the subject breathing slowly and deeply five times at five-minute intervals into a valved-Teflon sampling bulb (Markes Bio-VOC Sampler) containing a Tenax rapid passively sampling device in which the breath VOCs were captured. The subjects also completed a questionnaire concerning recent pollutant exposures and ingested food that may produce cofounding breath VOCs (the same type of questionnaire was previously used in a VOCs breast cancer detection study). The sampling device was analyzed by thermal desorption/gas chromatography/mass spectrometry (TD/GC/MS) (Markes International Ltd. ULTRA thermal desorber/Thermo Trace GC ULTRA/Thermo Trace DSQ mass spectrometer). The VOCs were identified from the GC/MS data at an off-campus laboratory by Dr. Charlene Bayer.

In addition to the breath samples, the following measures were taken on all visits: heart rate (HR), oxygen saturation (O_2 sat), blood pressure (BP), body mass (BM), urine color (U_C), urine specific gravity (U_{SG}), Visual Analog Scale of thirst (VAS), and urine osmolality (U_{osm}). O_2 sat and HR were measured using a pulse oximeter (Nonin Medical, Plymouth, MN). BP was measured using an automatic blood pressure monitor (OMRON Healthcare, Bannockburn, IL). BM was measured on the Pennsylvania 50 Scale (Wiggins Scale Co., Atlanta, GA). U_C was measure using a color chart previously described and validated[48]. U_{SG} was measured with a hand held refractometer (ATAGO). The Osmette micro osmometer (Precision Systems, Natick, MA) was used to determine U_{osm}. The VAS was a 100 mm scale with the anchors being “Not thirsty at all” and “As thirsty as possible”.

On the first day, body composition was measured to calculate total body water, assuming the change in body fat during those 3 days was negligible. Plasma osmolality

(P_{osm}) was measured in the last three visits from blood samples taken from the finger. Basal metabolic rate was also measured, using a metabolic hood calorimeter, during visits 1, 3, 4 and 5. In addition to the above measures, subjects completed questionnaires for 24-hour living on each visit. Subjects also maintained a food log following the afternoon visit on the first day. During the dehydration period, the subjects had to collect their urine for a 24-hour period.

2.2.2 Analysis methods

The outcome measure was hydration status, a binary variable with values of 1 or 0 indicating whether a subject was hydrated or dehydrated. The VOCs in each sample breath were used to classify the subjects as hydrated or dehydrated. Clinical variables such as sex, height, weight, blood pressure were used as potential covariates. Changes in body weight, blood and urine osmolality, and urine specific gravity samples were initially selected as gold standard biomarkers [5] to monitor the status of hydration/dehydration, to compare to the breath measures and to assess sensitivity and accuracy of categorical assignments (hydrated, dehydrated).

Mass spectrometry data is usually noisy and has more variables than the available number of samples ("small n, large p"). In addition, many VOCs for a particular subject are not observed (value is 0), which makes the data sparse. Thus, before doing any classification analysis, these issues must be addressed. To reduce or filter out the noise, some pre-processing of the data is necessary, such as standardization and/or normalization techniques. Then, data dimension reduction methods can be applied. Only samples from the last 3 visits were included for the analysis of hydration status classification. These were: euhydration, 16-hour dehydration and 24-hour dehydration.

The sequence of methods used in the analysis is depicted in Figure 2.1.

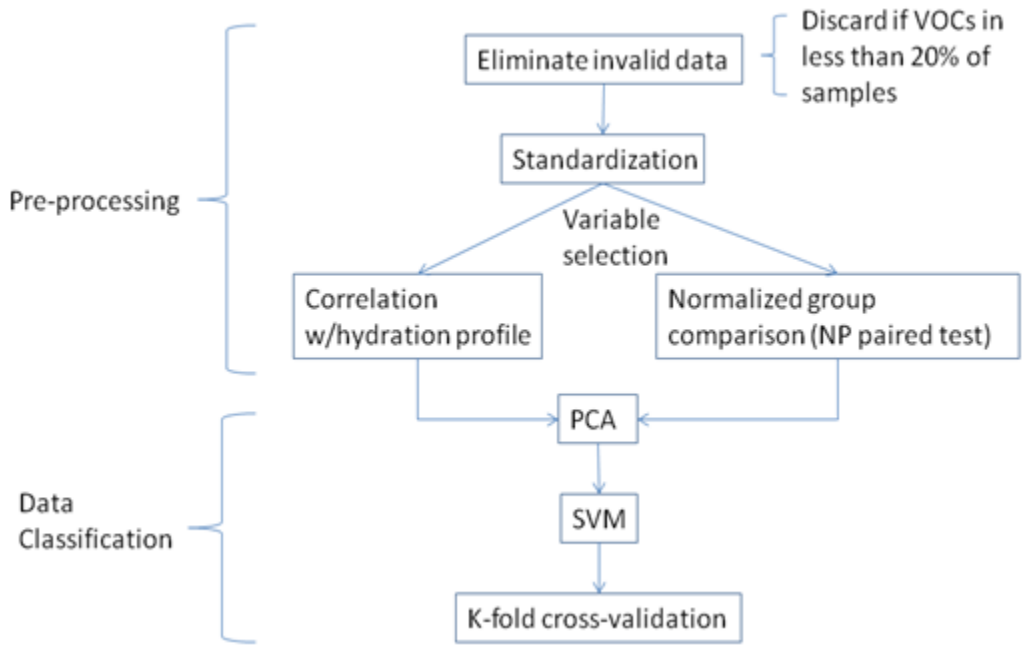

Figure 2.1 Diagram of the sequence and methods used in the analysis of VOCs

Pre-processing of data

A. Invalid Data/Missing Values

Usually, a percentage of the initially identified VOCs will appear in only a few observations, and thus, will not be useful for classifying the subjects. We chose 20% as the cutout value to eliminate "not useful" VOCs. Thus, if a VOC was present in less than 20% of the breath samples for each hydration group (i.e., it was present only in 7 or less subjects from the euhydration group, and in 7 or less subjects from the 16hr dehydration group and in 7 or less subjects from the 24hr dehydration group), then that VOC was eliminated.

B. Standardization

The remaining compounds were standardized across subjects to have mean 0 and variance 1.

C. Variable Selection or Filtering

To determine which VOCs were most useful in the classification of hydrated versus dehydrated subjects, we applied to two different variable selection methods:

1. *Normalized difference between two groups*: The nonparametric sign test for paired samples was used to compare the hydrated vs. the dehydrated group for each VOC. The VOCs were ranked and selected based on the absolute values of the statistical test. Classifiers were later evaluated for two groups of VOCs, one for p<=0.05 and another for p<=0.1
2. *Correlation with hydration profile*: An individualized hydration profile (the probability of a subject being dehydrated) was built using a logistic regression model with some of the hydration measures as predictor variables. The idea was to correlate the VOC profile with the hydration profile for each of the VOCs. A compound was selected if its absolute correlation value was higher than a specified value. Classification methods were evaluated on three groups: VOCs with corr >= 0.15, corr >= 0.18 and corr >= 02.

Subsequent classification algorithms were also used to compare the variable selection methods.

Data classification

To distill predictive summaries from the VOC data, Principal Component Analysis (PCA) was implemented. The PCA is a procedure that utilizes a linear transformation of data to form a set of variables that are uncorrelated. In addition, the

totality of these uncorrelated variables (the Principal Components, PCs) preserves the variance in the data; however, this variance is compressed in only a few PCs. This linear transformation is organized in such a way that the first principal component has the highest possible variance (i.e., accounts for as much of the variability in the data as possible), and each succeeding component has the largest possible variance, conditionally to being uncorrelated with all the preceding components.

In real applications, several principal components contain almost all variance and can describe the data/problem as accurately as the data itself. Even though PCA is a powerful data reduction technique, it is often impossible to attribute any physical meaning to any particular PC, since they are linear mixtures of all variables in the problem. Thus, application of PCA can be thought of as revealing the internal structure of the data in a manner that best accounts for the variance in the data. If a multivariate dataset is visualized as a set of coordinates in a high-dimensional data space (one axis per variable), PCs can supply the user with a lower-dimensional picture, a "footprint" of this multidimensional object when it is viewed from its "most informative viewpoint." This is done by using only a few principal components so that the dimensionality of the transformed data is drastically reduced.

A. Classification algorithm

Data was analyzed using MATLAB R2010b (ver. 7.11.0). Since the VOCs are likely to be correlated to each other, or at least grouped in clusters according to their chemical characteristics, we decided to further apply principal components in order to get orthogonal variables. For each subject, a set of PCs at euhydrated and dehydrated points was found. Usually, the 1st and 2nd principal components are used as classifying

descriptors; however, other pairs were also tested, since the discriminatory power may not be necessarily connected with the primary principal components [49].

As typical in machine learning theory, when we have supervised learning (i.e., when we know which PCs are euhydrated and which dehydrated), the PCs are divided into two sets: training and validation set. On basis of a training set, a classifier is developed and then assessed based on the validation data. Dudoit et al [50] found that simple classifiers such as linear discriminant analysis (DLDA) and next neighbor (NN) performed better than more sophisticated ones (e.g., classification trees) when applied to "large p, small n" problems. In a similar study to classify cancer based on MS spectra, the support vector machine (SVM) classifier had the lowest error rate when compared to other discrimination methods [51]. Other classification studies with MS data also showed the superiority of the SVM, which displayed a robust performance even when the number of variables was changed [52]. The SVM takes a set of training PCs, each marked as belonging to one of two categories (euhydrated/dehydrated), and builds a model that assigns new examples into one category or the other. The resulting SVM model is a representation of points in a plane, mapped so that the members of the separate classes are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. Unlike linear and quadratic classifiers, the boundaries between the classes could be curvilinear and adaptive to the data. These boundaries are established by using the so-called kernel functions that are sensitive and expressive.

The support vector machine classifier was selected for the classification analysis, and the VOCs were evaluated on different kernel functions.

B. Cross-validation

Classifiers generally perform better on training data that are used to construct them than the test data. Therefore, to make the estimates of performance measures even more reliable, researchers often use a K-fold cross-validation. The original MS (training) data set is randomly split into K sets of about equal size. For each $1 \leq k \leq K$, the k^{th} part is regarded as a test and its complement as the training set. The classifier is built using this (artificially created) training set and its performance measures are computed using the partition that was left out (artificially created test set). When all K classifiers have been built and tested, the estimated error rates are averaged across all partitions [53]. Because of the small number of samples, we selected K=number of subjects. The two dominant components served as classifying descriptors. The discriminatory boundaries were created using the training sets and correct classification sets assessed by the validation sets. This was repeated 20 times.

Comparison with hydration markers

Finally, we compared the performance of our classifier against two well-known markers of hydration: urine osmolality (U_{osm}) and plasma osmolality (P_{osm}) [2, 9, 17]

2.3 Results

The final de-identified dataset consisted of 46 subjects. There were 23 males and 23 females, with an average age of 23.7 yr., initial average weight of 152 lbs and height of 172.4 cm. Men were heavier (169.3 lbs) compared to women (135.1 lbs) and taller by 12.8 cm. The mean values of the hydration measures for each of the 5 visits are presented in Table 2.1.

After some preliminary analysis and input from Dr. Millard-Stafford, five variables were excluded from further analysis: Usg, Ucol, 24-hr Uvol, 24-hr Usg and thirst rating. The remaining variables were: Sex, Uosm, Posm, TBW and VAS.

Table 2.1 Hydration measures for all 5 visits

Measure	Mean V1 (St. Dev)	Mean V2 (St. Dev)	Mean V3 (St. Dev)	Mean V4 (St. Dev)	Mean V5 (St. Dev)
Weight (lbs)	152.0 (31.8)	151.9 (31.7)	152.1 (31.9)	150.0 (31.5)	148.9 (31.2)
Uosm	767.3 (288.9)	736.6 (238.0)	381.7 (176.9)	949.0 (101.2)	1033.5 (96.1)
Usg	1.021 (0.008)	1.019 (0.006)	1.010 (0.005)	1.027 (0.003)	1.027 (0.003)
Ucolor	5.1 (1.4)	4.1 (1.4	2.5 (1.4)	6.1 (0.7)	5.9 (0.8)
Total Body Water	94.2 (20.9)	94.2 (20.9)	94.3(21.1)	92.7 (20.7)	91.9 (20.5)
Posm	--	--	284 (4.4)	290.6 (5)	292.8 (5.2)
VAS	3.6 (2)	3.4 (2.2)	2.3 (1.7)	6.4 (1.4)	7.3 (1.8)
Thirst Rating	3.5 (1.5)	3.7 (1.9)	2.7 (1.6)	6.4 (1.4)	7.5 (1.4)
24-hr Uvol	--	--	--	--	846.9 (348)
24-hr Usg	--	--	--	--	1.022 (0.004)

Breath samples from 10 subjects were found to be contaminated and discarded. VOCs for the remaining 36 subjects (22 females and 14 males) were assessed. A total of 331 VOCs were detected in each subject's breath sample. Measurements on each subject included in the study had two values: VOCs at the point of euhydration and at 24-hour dehydration, for a total of 36 pairs of measurements.

The first step in the reduction of VOCs eliminated 242 compounds. As mentioned in the methods section, we considered "invalid" data those compounds that appeared in so few samples that would not be of any value during the classification. A VOC was excluded if it appeared in less than 29 subjects, i.e., VOCs that only appear in 7 samples

or less for each of the 3 hydration groups. The second step in reducing the number of VOCs involved filtering the data in two ways:

1. Correlating each VOC with a template obtained from the hydration measures (hydration profile)
2. Using a statistical test to identify VOCs that discriminated between hydration groups

2.3.1 Selection of VOCs using Hydration Profile

To obtain the hydration profile for each subject, several logistic regression models were evaluated (Table 2.2). Sex was a significant predictor in most models, except when P_{osm} was included.

Table 2.2 Logistic Regression Models for hydration measures

	Predictors						Cox &	Nagelkerke	% Correct. Class.	
Model	Sex	Uosm	Posm	TBW	VAS	-2L	Snell R2	R2	Training	Test
1	x	x		x	x	17.353	0.698	0.931	95.7	91.3
2	x	x		x		24.045	0.675	0.900	94.6	87.0
3			x		x	49.174	0.573	0.764	90.2	87.0
4	x			x	x	51.061	0.565	0.753	88.0	80.4
5	x	x				97.264	0.280	0.374	75.0	100.0

In addition, only euhydration and 24-hr dehydration were considered in the profile. This was due to observing that the values of VOCs were unstable at 16-hr dehydration. Unlike the other hydration measures, the VOCs values at 16-hr dehydration did not always fall between euhydration and 24-hr dehydration. Figure 2.2 shows the hydration profile for the 36 subjects, 2 samples for each one.

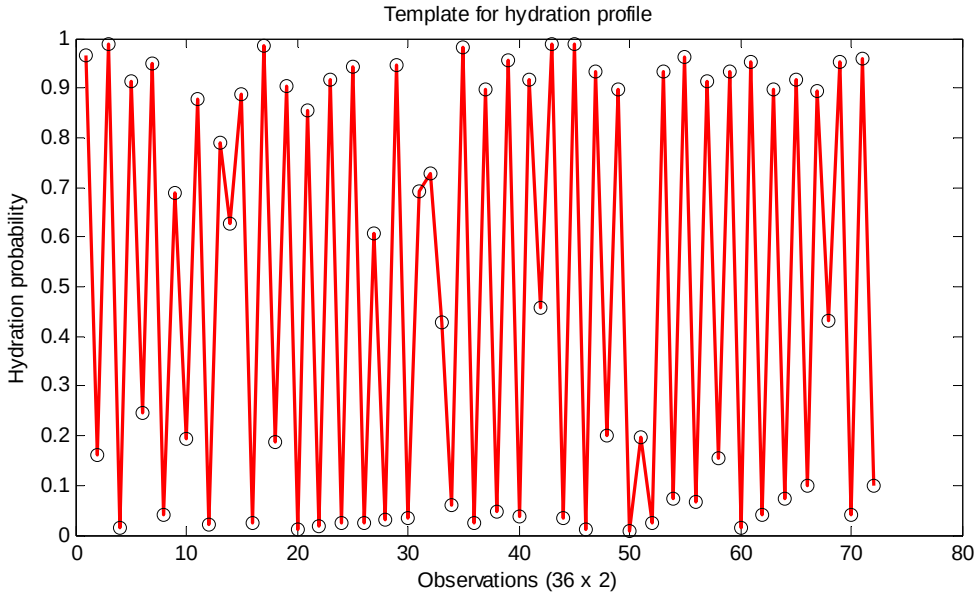

Figure 2.2 Hydration Status Profile for the model with P_{osm} and VAS

All 89 VOCs were correlated with the best three hydration profile models. No strong correlations were found between any particular compound and the hydration profiles. Most of the compounds had correlations values in the [-0.15, 0.15] range. Of the hydration models, we selected the third one (Posm, VAS) as the hydration profile template because it presented the highest absolute correlation value (Table 2.3).

Table 2.3 Results of correlations between VOCs and different hydration profiles

Model	Predictors in model	Correlation range	# VOCs (corr > 0.2)	# VOCs (corr > 0.18)	#VOCs (corr > 0.15)
1	Sex, U_{osm}, TBW, VAS	[-0.24, 0.23]	5	8	17
2	Sex, U_{osm}, TBW	[-0.25, 0.24]	4	6	15
3	**P_{osm}, VAS**	**[-0.28, 0.20]**	**2**	**5**	**13**

The VOCs with the largest absolute correlation values were selected, and several classification algorithms were evaluated on different number of VOCs (grouped by their

correlation values). Principal component analysis was performed on the groups of selected VOCs previous to applying the classification algorithms. A linear support vector machine classifier was applied to different pairs of principal components. The highest correct classification rate was achieved when 13 VOCs (corr > 0.15) were used (Table 2.4). The following kernel functions were further evaluated on the transformed VOCs: Linear, Quadratic, Gaussian radial basis (sigma=1 and 2) and Polynomial. The results presented in Table 2.5 show that the linear kernel provided the highest correct classification rate. The 1st and the 3rd principal component were the ones that best classified the samples (Figure 2.3).

Table 2.4 List of VOCs with abs(corr)>0.15

VOC	Correlation value
Triacetin	-0.2787
Limonene	0.2018
2,3-dihydrofuran	-0.1958
Methenamine	-0.1908
2-butoxyethanol	0.1846
Decane	0.1785
Heptacosane	-0.1781
Pentacosane	-0.1617
Thiourea	-0.1597
2-methylfuran	0.1577
Cyclohexane	0.1563
Dodecane	0.1541
N-ethylethanamine	-0.1515

Table 2.5 Results from SVM for different kernel functions (K-fold, 30 repetitions)

Cutout for correlation	# of VOCs	Kernel function	PCs w/ best CR	Correct Rate (CR)	Sensitivity	Specificity
0.2	2	Linear	1,2	0.6503	0.6511	0.6495
0.18	5	Linear	1,2	0.6965	0.7037	0.6893
0.15	**13**	**Linear**	**1, 3**	**0.7389**	**0.7242**	**0.7536**
0.15	13	Quadratic	1,3	0.6831	0.6375	0.7286
0.15	13	RBF (sigma=1)	6,9	0.6451	0.6393	0.6508
0.15	13	RBF (sigma=2)	1,3	0.6992	0.6448	0.7536
0.15	13	Polynomial (order=3)	1,3	0.6524	0.5213	0.7835

Highlighted row is the kernel function with the highest correct classification rate

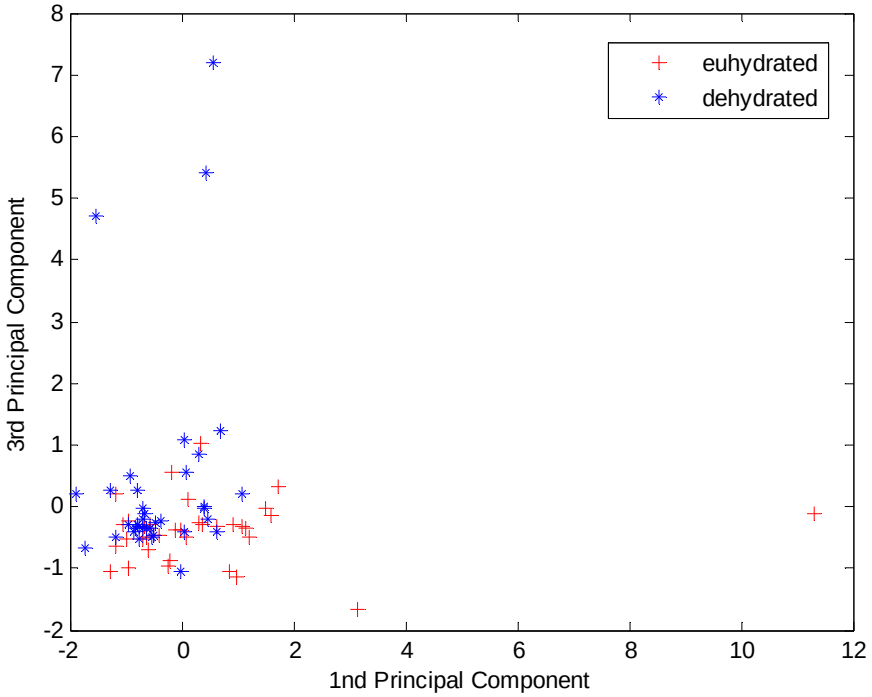

Figure 2.3 1st and 3rd Principal Components for the correlated VOCs

The performance of the classifier was evaluated using the K-fold cross-validation method. The classifier was trained and evaluated K=10 times, and this was further repeated 30 times. The estimated error rates were averaged across all partitions and repetitions. In this case, the specificity value indicated the rate of dehydrated subjects

correctly classified as being dehydrated. In Figure 2.4, euhydrated and dehydrated subjects are indicated by 0 and 1, respectively.

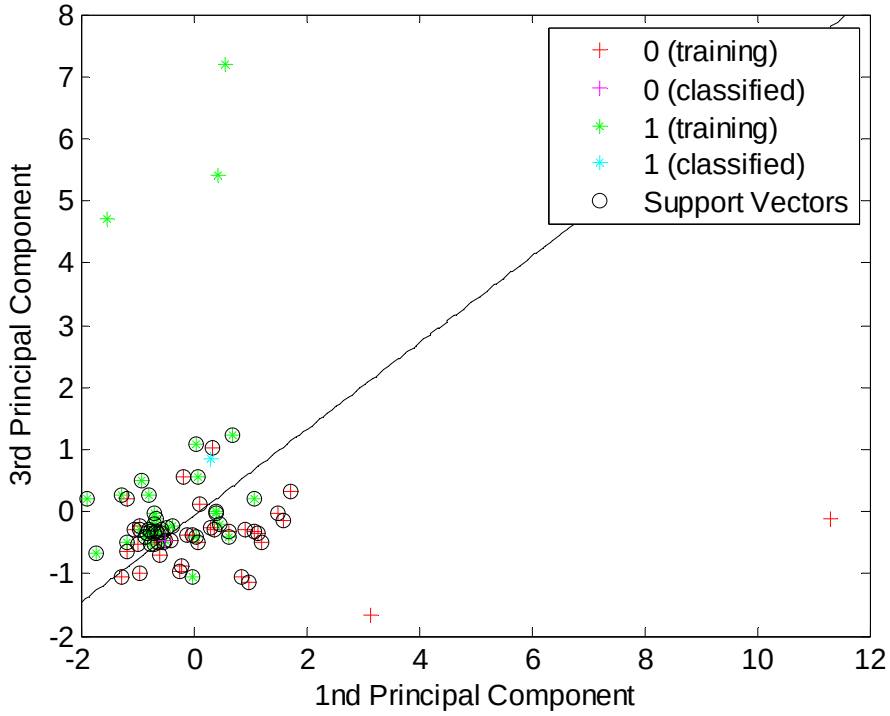

Figure 2.4 Classification with linear Support Vector Machine classifier

2.3.2 Classification of VOCs selected using statistical tests

Another method to reduce the number of VOCs used for classification, without using hydration measures, consists in performing a preliminary selection of VOCs based on the ratio of their between-group to within-group sums of squares (BW ratio). The larger the ratio, the more likely the compound will be relevant to the classification. Since this was a repeated measures design, we used a nonlinear test for paired samples to compare between hydration groups (euhydration vs. 24hr dehydration). The VOCs with the largest BW ratios were selected and used in the classifiers. Before applying the classifier, the selected VOCs were transformed into their principal components.

Previously, while building the hydration profile (refer to Table 2.2), we found that gender was significant in almost all the models. Thus, we decided to also include it in the classifier and the variable gender was then added to the selected VOCs when calculating the PCs. The classifiers were evaluated for two groups of VOCs using the following cutout values, p=0.05 and p=0.1. The classified compounds are listed in Table 2.6. The first 4 compounds have p <= 0.05. When all 9 VOCs were used, the 6^{th} and the 9^{th} PCs were selected for the classification. For the group of 4 VOCs, the 2^{nd} and the 3^{rd} PCs obtained the highest classification rates. The results in Table 2.7 indicate that the linear SVM classifier applied to the group of 4 VOCs generated the highest correct classification rate. Also, the results from adding the gender vector are slightly higher than when not using it.

Table 2.6 Selected VOCs with p <= 0.1

VOC name	p-value
triacetin	0.0127
limonene	0.0288
hexamethylcyclotrisiloxane	0.0347
2-Butoxyethanol	0.0352
toluene	0.0501
(Z)-1-(methylthio)-1-propene	0.0574
1-nonene	0.0923
allyl methyl sulfide	0.0963
α-Pinene	0.0963

Table 2.7 Results for evaluating different SVM classifiers on the transformed VOCs

PC	Kernel function	Correct Rate	Sensitivity	Specificity
p <= 0.1 (9 VOCs)				
6,9	Linear	0.6404	0.6131	0.6678
p <= 0.05 (4 VOCs)				
2,3*	Linear	0.6646	0.5523	0.7768
2,3	**Linear**	**0.6671**	**0.5567**	**0.7776**
2,3	Quadratic	0.6061	0.4988	0.7135
2,3	RBF (sigma=2)	0.6523	0.5567	0.7478
2,3	RBF (sigma=1)	0.6211	0.5258	0.7165
2,3	Polynomial	0.5933	0.4389	0.7478

* In this case, the female vector was not added to the 4 VOCs

The 2nd and the 3rd principal component were the ones that best classified the samples (Figure 2.5)

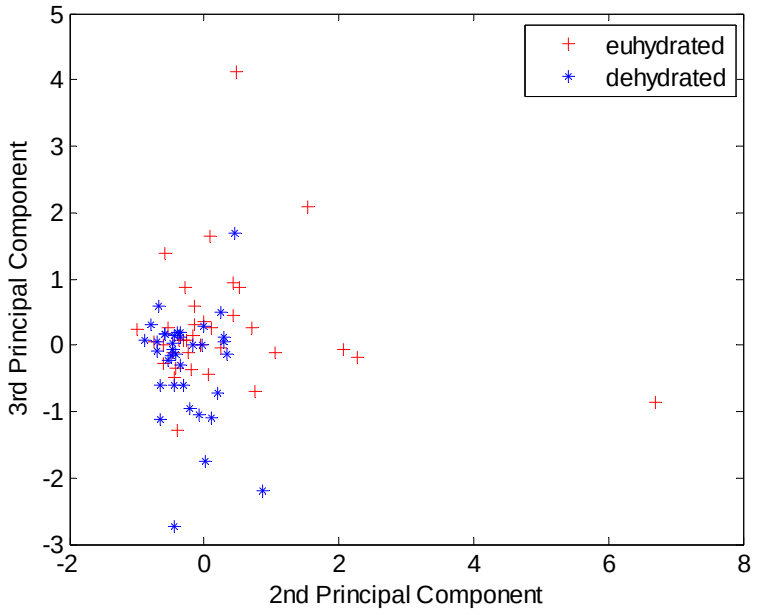

Figure 2.5 VOCs with p<=0.05 represented by their 2nd and 3rd PCs

The performance of the classifier was evaluated using the K-fold cross-validation method. The classifier was trained and evaluated K=10 times, and this was further repeated 30 times. The estimated error rates were averaged across all partitions and

repetitions. In this case, the specificity value indicated the rate of dehydrated subjects correctly classified as being dehydrated. In Figure 2.6, euhydrated and dehydrated subjects are indicated by 0 and 1, respectively.

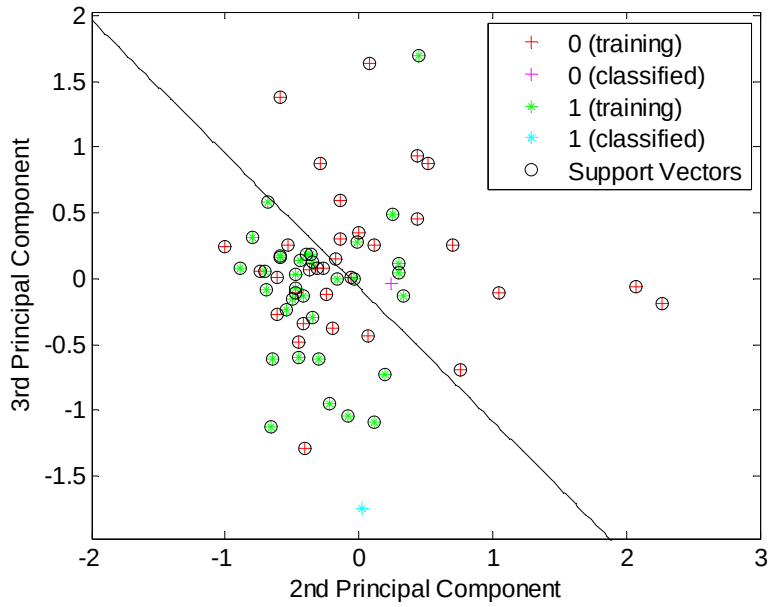

Figure 2.6 Results from linear SVM classifier using the 2nd and 3rd PC

2.3.3 Comparison of the breath classifier with other hydration measures

The performance of the breath classifier was compared to two well known hydration measures: U_{osm} and P_{osm}. Several logistic regression models to predict hydration and 24hr dehydration were built. The classification rates of all the models are presented in (Table 2.8). The best classification rate is achieved by U_{osm}, which correctly predicted almost all subjects (99.3%). P_{osm} only achieved 81.2%, however, when VAS was added to the model, the classification rate was 92.8%.

Table 2.8 Classification tables for several measures of hydration

Measure: Uosm

Observed		Predicted		
		HydStatus		Percentage Correct
		Dehydrated	Hydrated	
HydStatus	Dehydrated	91	1	98.9
	Hydrated	0	46	100.0
Overall Percentage				99.3

Measure: Posm

Observed		Predicted		
		HydStatus		Percentage Correct
		Dehydrated	Hydrated	
HydStatus	Dehydrated	79	13	85.9
	Hydrated	13	33	71.7
Overall Percentage				81.2

Measure: Posm + Sex

Observed		Predicted		
		HydStatus		Percentage Correct
		Dehydrated	Hydrated	
HydStatus	Dehydrated	86	6	93.5
	Hydrated	13	33	71.7
Overall Percentage				86.2

Measure: Posm + MC

Observed		Predicted		
		HydStatus		Percentage Correct
		Dehydrated	Hydrated	
HydStatus	Dehydrated	85	7	92.4
	Hydrated	10	36	78.3
Overall Percentage				87.7

Measure: Posm + VAS

Observed		Predicted		
		HydStatus		Percentage Correct
		Dehydrated	Hydrated	
HydStatus	Dehydrated	88	4	95.7
	Hydrated	6	40	87.0
Overall Percentage				92.8

MC: Menstrual Cycle; VAS: Visual Analog Scale for perception of thirst

The correct classification rate for breath analysis, when selecting VOCs with the hydration profile, was 73.89%, which was lower than any of the rates in Table 2.8.

From these results, it is clear that breath analysis is still not a strong marker of hydration status, and other more established hydration markers have superior performance.

A few limitations in this study were:

- In this type of study design, it is difficult to obtain a large enough representative sample of the overall population. It usually represents only a group of the total population (in this case, healthy young people ages 18 to 40).
- Although the initial number of subjects is 46, the number of available breath samples was reduced by 10 due to contamination of the samples or malfunction of the equipment.
- Mass spectrometry (MS) has the potential to identify more sensitive biomarkers of a health condition than current ones. However, the process is very sensitive to changes in the protocol of sample and spectra collection [53]. The MS data provided to us was already preprocessed; therefore, our results were restricted to the quality of the preprocessing steps.
- During the 8 months of data collection, two different MS equipment were used for the analysis of the samples. The machines did not identify the same set of VOCs for the breath samples, creating additional variability in the process.

2.4 Conclusions

Exhaled breath samples from 36 subjects were evaluated. Two measurements from each subject were analyzed: euhydration and 24-hour dehydration. Data dimension and classification methods were applied to the 331 VOCs found in each breath sample. When the sign test was used to select relevant VOCs, the highest correct classification rate achieved was 66.71%. When VOCs were selected by correlating them to the hydration profile, the classification rate was 73.89%. Although these results are better than flipping a coin (50%), they are still inferior when compared to other used marker of hydration such as U_{osm} and P_{osm}.

Breath analysis presents large variability in the number and concentration of VOCs [32], thus, until a true criterion in hydration is clearly established for a single point in time measure (as compared to following changes from known euhydration to dehydration), many more samples would need to be obtained (with substantial investment in the technology). Additionally, a better understanding of the specific VOCs involved in the physiological changes due to dehydration is very much required. For instance, in our study, most of the VOCs selected for classification were compounds to which subjects may have been exposed from the environment - such as limonene (emitted from building materials), 2-butoxyethanol (found in many cleaning products), decane, heptacosane, heptacosane, dodecane (alkaids found in combustibles), and triacetin (found in cosmetics and personal care products), and their role in the dehydration process is not clear.

Although our results did not provide an improvement on other hydration measures, application of breath analysis to areas such as classification of diseases has

been more successful [32, 42, 44]. In the topic of hydration classification, future research could include:

- Evaluation of the validity of using fingerstick P_{osm} as a less invasive measure (compared to venipuncture) of hydration status.
- Analysis of the VOC signature of the other dehydration group: 16-hours dehydrated and evaluate classification models to discriminate against the euhydrated and the 24hr dehydrated groups.
- Identify from the literature and field experts, the VOCs that can potentially describe the physiological changes that occur during dehydration.

CHAPTER 3 : IMPACT OF EMR ON MEDICAL ERRORS AND ADVERSE DRUG EVENTS

3.1 Literature Review

3.1.1 Patient safety, medication errors and adverse drug events

According to the 2007 report Preventing Medication Errors by the Institute of Medicine, medication errors are the most common medical errors, and the report *"estimated that on average, a hospital patient is subject to at least one medication error per day, with considerable variation in error rates across facilities"* [54]. These findings are further supported by other studies identifying the high incidence of medication errors within healthcare facilities [55, 56]. Although not all errors lead to injury or death, the report found that medication errors injure at least 1.5 million people per year. Medical errors can happen at different stages in the medication process: prescribing, filling, dispensing, or administering; however, most of them occur during prescribing [56, 57]. Therefore, the implementation and utilization of an electronic prescribing system would greatly help reduce the frequency of medication errors [58].

Using the definition of the National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP), a medication error is "*any preventable event that may cause or lead to inappropriate medication use or patient harm while the medication is in the control of the health care professional, patient, or consumer. Such events may be related to professional practice, health care products, procedures, and systems, including prescribing; order communication; product labeling, packaging, and*

nomenclature; compounding; dispensing; distribution; administration; education; monitoring; and use". Similarly, and adverse drug event (ADE) is defined as "*an injury resulting from medical intervention related to a drug"* [56]. ADEs may or may not result from medication errors, and studies show that few ADEs actually result from them [56], occurring most often during the ordering stage [59]. As in the case with medication errors, health information technologies (HITs) such as computerized physician order entry (CPOE) systems can help prevent many of these ADEs. A preventable ADE is an injury that is the result of an error at any stage in the medication use. A non-preventable ADE, a.k.a., adverse drug reaction (ADR), is an injury due to a medication, but in this case there is no error in the medication process [60].

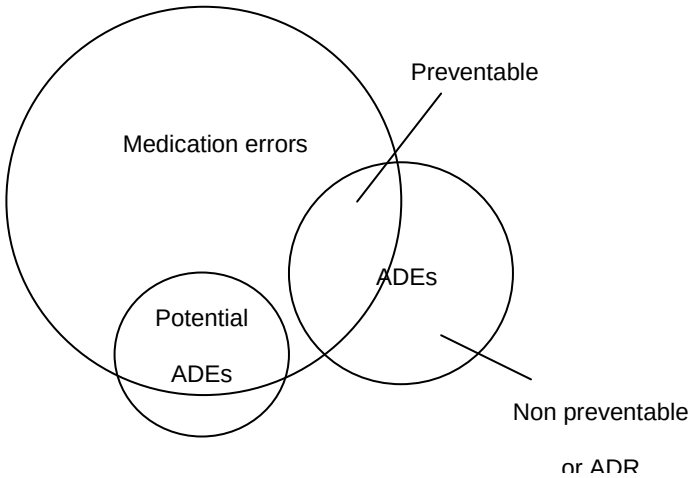

Figure 3.1 Relationship between medication errors and adverse drug events [60]

There is variability in the number of medication errors and ADEs reported by healthcare organizations, mainly due to the lack of standardization in the definitions and in the methods used for the detection of events [61]. Differences across hospitals units

have been identified in the rate of ADEs [62], as well as in the rate of reported medication errors [63]. Due to the complexity of patient cases and medications in the intensive care units (ICUs), medical errors and adverse events have been reportedly higher than in general care units [63]. There are also differences regarding patient populations. Children are at greater risk than adults for medication errors [64]. Children in particular are more susceptible to serious medication errors due to weight-based dosing, off-label drug usage and preparation, limited ability to withstand a dosing error and a limited ability to communicate with healthcare professionals when an error might occur or has occurred [65, 66]. In a study in two academic pediatric hospitals, it was found that errors occurred at a rate of 5.7 errors per 100 orders, with 79% of these errors occurring at the ordering stage [66]. Errors with potential to cause harm were three times more likely to occur in pediatric inpatients compared with adults [56, 66]. The risk for adverse events is even higher in pediatric ICUs [67]. Other factors that put patients at higher risk for ADEs and medication errors are increased drug exposure and hospital stay [68, 69].

3.1.2 HIT and patient safety

One of the recommendations of the 2007 IOM report to improve patient safety was for healthcare organizations to have electronic prescribing in place by 2010 [54]. A recent national survey on US hospitals from 2011 found that 26.6% of hospitals had at least a “basic” electronic health record (EHR) system, while 8.7% had a “comprehensive” system. From a previously developed definition, a basic system has the following deployed technologies in at least one hospital unit: computerized systems for patient demographics, physician notes, nursing assessments, patient problem lists, laboratory and

radiologic reports, diagnostic test results, and order entry for medications. A comprehensive EHR system must include all of the functions that a basic system can perform and fourteen additional functions deployed in all major hospital units [70].

Since most medication errors occur during the ordering and prescribing process, electronic prescribing has the potential to greatly reduce the risk of medication errors and ADEs. Electronic prescribing is provided by Computerized Physician Order Entry (CPOE) systems, which refers to computer-based systems that automate the medication ordering process [71, 72]. CPOE systems can ensure legible and complete orders and assist the physician during ordering by providing evidence-based decision support tools [73] such as checking drug-drug interactions, drug allergies, or drug-laboratory interactions [74] . For example, the system can check for age specific dosing regimens and doses above or below the usual range, display alerts to the user if the current laboratory values indicate that the drug would be inappropriate for a certain patient. In addition, data collected from such systems can be used to automatically detect signals associated with an adverse reaction [75]. Although the evidence regarding the effectiveness of CPOE to reduce prescribing errors is still modest [76], several studies have demonstrated that CPOE systems help decrease the frequency of errors and ADEs in both the adult and the pediatric settings [77-81], as long as they also include the adequate decision support mechanisms [82, 83]. CPOE would be even more effective if it were connected to other computerized system [84] that would prevent the remaining routes for errors (e.g., during the medication administration process) [85]. An inpatient EHR with CPOE in a teaching hospital decreased medication errors per 1000 hospital days from

17.9 to 15.4 and the percentage of medication events that were medication errors from 66.5% to 55.2% [86].

Transcription errors can be reduced using electronic medication administration record (eMAR) systems. The main functions of these systems are to organize medication administration schedules and ensure timely medication administration [87]. In a study by Mekhjian, CPOE combined with eMAR completely eliminated physician and nursing transcription errors [88].

Other systems that can reduce medication errors and ADEs are pharmacy-based computer systems and smart pumps. Although their usefulness might be limited to interventions after an order has been written, pharmacy-based computer systems help prevent fundamental ADEs such as drug-allergy and drug-drug interactions. Smart pump technologies provide clinicians with safeguards for proper dosing; however, to detect and avoid medication errors, smart pumps have to be properly implemented and clinicians have to be properly trained to use them [89].

The implementation strategy of an IT system is also an important of preventing medication errors [74]. For instance, a meta-analysis study that evaluated the effectiveness of CPOE systems found that its implementation process may be associated with adverse outcomes and can decide the success or failure of the system [90]. A recent study review highlighted the “human element” as a critical component to the implementation of HITs, finding a strong association between provider satisfaction and negative outcomes [91]. It is worth mentioning that despite the numerous studies on the positive impact of HIT, a few of them have also described the potential for new kind of

medication errors, not only because computer interfaces but also because of human errors [92, 93].

3.1.3 Analysis of medication errors and adverse drug events

The outcome of interest when evaluating ADES and medication errors is usually some type of count data, e.g., frequency, incidence, or rate of events. Most studies found in the literature that analyzed medication errors and/or ADEs in hospitals used univariate analysis such as t-test, one-way ANOVA or χ^2 test [72, 94]. These techniques are valid to evaluate associations of independent variables to outcomes variables; however they don't account for the effect of covariates. A few studies used logistic regression to identify the risk factors associated to ADEs and medication errors [63, 95-98], but these models only predicted the absence or presence of the event.

Methods for modeling count data

There are several methods that can used to model count data. The most commonly used method to model count data is the Poisson regression [99]. If the data presents overdispersion or a high frequency of zero counts [100], it would violate the Poisson assumption of variance and mean equality. In those cases, applying a Poisson model would underestimate the observed dispersion, resulting in improper estimates of the standard errors and increasing the type I error. One approach for this problem is to use the negative binomial, a more flexible model than the Poisson, which allows the variance to be greater than the mean. An alternative to the negative binomial, for cases when the data presents a large frequency of extra-zeros, is to use either the zero-inflated or hurdle models [101, 102]. The options are the zero-inflated Poisson (ZIP), zero-inflated negative binomial (ZINB), and similarly, the hurdle Poisson and the hurdle negative binomial.

Zero-inflated models assume that the population consists of two different groups, with the zeros coming from both of them. A description of the aforementioned models follows:

a. Poisson regression is a form of a generalized model in which the response variable y_i is modeled as having a Poisson distribution, and the explanatory variables x_i' (vector of linearly independent regressors) determine the mean of the response variable, i.e., μ_i. The Poisson regression model is defined as follows [103]:

   Density function of y_i given $\mathbf{x}_i$:

$$f(y_i|\mathbf{x}_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!},\ y_i = 0, 1, 2, \ldots \quad (1)$$

   To ensure that $\mu_i > 0$, the mean parameter is parameterized using the log link function:

$$\mu_i = \exp(\mathbf{x}_i'\beta) \quad (2)$$

b. Negative Binomial: The most common form of the negative binomial model is the one with variance function $\mu + \alpha\mu 2$ and density function:

$$f(y|\mu, \alpha) = \frac{\tau(y+\alpha^{-1})}{\tau(y+1)\tau(\alpha^{-1})}\left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu}\right)^{\alpha^{-1}}\left(\frac{\mu}{\alpha^{-1}+\mu}\right)^{y}, \quad \alpha \geq 0, \quad y = 0,1,2, \ldots \quad (3)$$

   When $\alpha = 0$, this reduces to Poisson.

c. Zero-inflated count models assume that the population consists of two distinct groups, with zero counts generated from both groups. The ZIP model for the response y_i can be represented as [101]:

$$y_i \sim 0 \qquad \text{with probability } \varphi_i \quad (4a)$$

$$y_i \sim \text{Poisson}(\mu_i) \qquad \text{with probability } (1-\varphi_i) \quad i = 1, \ldots, n \quad (4b)$$

   so that:

$$\Pr[y_i = 0] = \varphi_i + (1 - \varphi_i)e^{-\mu_i}, \tag{5a}$$

$$\Pr[y_i = r] = (1 - \varphi_i)\frac{e^{-\mu_i}\mu_i^r}{r!}, \qquad r > 0, \tag{5b}$$

where

$$\text{logit}(\varphi_i) = \gamma' w_i \tag{6}$$

and

$$\log(\mu_i) = \beta' x_i \tag{7}$$

A characteristic of this model is that the effects of the covariates can be examined simultaneously in both components. The covariates are not necessarily the same for both components, and two different set of covariate vectors are allowed in these models. For the ZINB, the Poisson distribution is replaced by the negative binomial in the ZIP model.

d. The Hurdle model can also be interpreted as a two-part model, with the first part being a binary outcome model, and the second part a truncated count model. This is useful for the cases in which the data may come from two sources, thus partitioning the population into subpopulations [103]. If the hurdle is set at zero, it would result in a group with zero elements, and a second one with positive counts. This model is defined as follows:

$$\Pr[y_i = 0] = f_1(0) \tag{8a}$$

$$\Pr[y_i = j] = \frac{1 - f_1(0)}{1 - f_2(0)} f_2(y_i) = \varphi f_2(y_i), \qquad j > 0 \tag{8b}$$

where $\varphi = (1 - f_1(0))/(1 - f_2(0))$ can be interpreted as the probability of crossing the hurdle.

Evaluation and selection of count data models

The models described above can be compared and evaluated in different ways:

- To identify overdispersion in the data, several tests can be applied. An indication of the magnitude of overdispersion can be obtained by simply comparing the sample mean and sample variance of the dependent variable. Application of the Poisson regression decreases some of the conditional variance of the dependent variable. However, if the sample variance is more than twice the sample mean, then the data is likely to remain overdispersed after the inclusion of regressors [76].
- Comparison of nested models such as Poisson vs. Negative Binomial or ZIP vs. ZINB can be compared using likelihood ratio, the Wald and the Score tests.
- Similarly, methods to compare nonnested models (Poisson vs. ZIP, NB vs. ZINB) include the Akaike Information Criteria (AIC), the Bayesian Information Criteria (BIC) and the Vuong test.

$$AIC = -2\log(\mathrm{L}) + 2p, \quad \text{where } p \text{ is the number of parameters} \tag{9}$$

$$BIC = -2\log(L) + \log(n)\,p, \quad \text{where } p \text{ and } n \text{ are, respectively, the number of parameters and observations} \tag{10}$$

- Finally, goodness-of-fit tests can also be used to select models based on how well the predicted values compare with the observed ones.

3.1.4 Summary of Key Research Needs and Specific Aims

From the literature review, we extracted the following points:

1. Most medical errors and adverse drug events occur during the prescription and ordering processes.

2. Because of 1), healthcare information technologies such as eMAR, CPOE and EMR have the potential to greatly reduce the number of medical errors and adverse drug events.
3. Although implementation of HIT systems is growing, only a small percentage of hospitals have implemented "comprehensive" electronic health systems (8.7% in the 2011 AHA survey).
4. The mere deployment of information technology systems in a healthcare organization does not guarantee the improvement of patient safety and healthcare quality.
5. New types of errors are bound to appear with the implementation of HIT systems.
6. Rates of medication errors and adverse events are not the same across a healthcare organization, so there is a need to evaluate them according to those differences.
7. Healthcare outcome measures like medication errors and ADEs are count data, so linear regression methods are not appropriate. Likewise, because of the high number of zero events, simple Poisson models might not always be a good fit.

This chapter attempts to contribute to the knowledge base of healthcare in two ways. First, it increases the knowledge base of implementing HITs by developing a better understanding of the impact of an incremental EMR implementation on patient safety, measured by the change in the rate of medical errors and adverse drug events. We focused on the pediatric inpatient setting because the medication administration process in children is more complicated [66], and thus, the occurrence of a medication error is more critical. Additionally, it was also of interest to examine the impact of HITs on reducing the differences between hospital types and areas.

The second main objective of this chapter is to identify and recommend the most adequate model for the analysis of count data in healthcare, in particular of the occurrence of medication errors, which usually presents a high number of zeros. The occurrence of events or incidents in healthcare settings exhibits some particular characteristics that must be considered when selecting a distribution that will fit the data. For those cases, logistic regression models do not suffice. In this study, we discuss the application and evaluation of six alternative nonlinear regression models to describe the frequency of medication errors and adverse events.

3.2 Methodology

This chapter examined the impact of an electronic medical record system implementation on the rate of medical errors and ADEs in two campuses of a pediatric healthcare system. Details regarding the study context, population and (a priori) data collection are described below, as well as the statistical analysis methods used.

3.2.1 Study Context

This was a retrospective longitudinal study of patients at two inpatient facilities, one academic and one community-based, in a pediatric health care system. Children's Healthcare of Atlanta (CHOA) implemented a series of electronic medical record (EMR) components over several years to improve patient safety and quality as well as increase efficiency of all operations. The system implementation occurred over four years (2004-2008). The implementation stages that were included in this study, in order of implementation are:

1. Inpatient Pharmacy System (Jan 2005)
2. Electronic Medication Administration Record and Clerk Order Entry (Nov 2005)

3. Nursing & Ancillary Documentation (May 2007)
4. Physician Documentation (Aug 2007)
5. Computerized Provider Order Entry System (June 2008)

This phased implementation approach allowed for the introduction and mastery of each EMR component separately, and for the separate analysis and reporting of each component and its contribution to improving patient safety and quality of care.

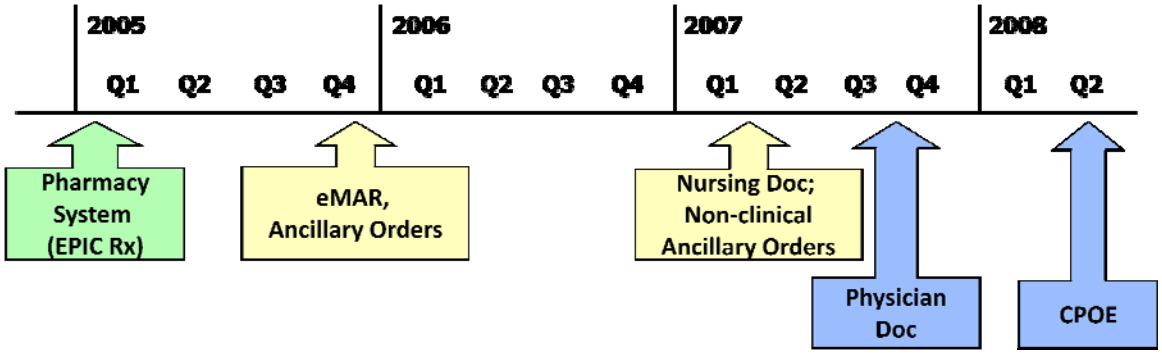

Figure 3.2: EMR Implementation Stages

Data Collection

Data collection for this study happened at six periods of time: the first, baseline, two months before the implementation of the first EMR component (EPIC Rx) and then, 6 to 8 months after the implementation of each additional EMR functionality. Since we were particularly interested in the impact of the EMR components involving the physicians, a baseline measurement was also taken before their implementation (T3).

A randomly selected list of patients was generated for each phase of the data collection. Patients were eligible for inclusion if they belong to either the general or the intensive care area, and had a minimum of 2 day admission.

Table 3.1 Data collection periods

Time Period	Phase
T0 (Nov 04 – Feb 05)	Prior to EMR implementation
T1 (Apr 05 – Jun 05)	Post EPIC Pharmacy System (EpicRX)
T2 (May 06 – Jun 06)	Post eMAR / Smart Pumps
T3 (Feb 07 – Mar 07)	Baseline measure Pre CPOE / Safety initiatives
T4 (Sep 07 – Nov 07)	Post Nursing & Physician Documentation
T5 (Sep 08 – Oct 08)	Post CPOE

Identification and classification of medication errors and adverse drug events

Retrospective reviews of the medical charts were conducted by researchers from the pharmacy department at CHOA. A trigger chart tool was used to assist in identifying possible medication errors and adverse drug events. This method was selected because several studies have demonstrated that it is more effective than voluntarily reporting of incidents to identify ADEs in pediatric hospitals [104, 105]. The tool was created from the Institute for Healthcare Improvement (IHI) trigger tool for measuring adverse drug events and a list from the Institute of Safe Medication Practice (ISMP). It collected baseline demographics such as date of birth, admission, discharge and total number of doses a patient received during a maximum of 14 days. The data collection tool directed the reviewer to five specific parts of the chart: 1) discharge summary, 2) orders, 3) anesthesia records, 4) medication administration record and 5) ancillary nursing notes. A list of common antidotes and medications used for treating adverse drug events were evaluated in the trigger tool. If a potential error or adverse event was identified, a brief

description of the incident was documented. The reviewer would then score the incident using a list of harm categories (see Table 3.2) based on the system for classifying the medication errors by the National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP). The data collection sheets were transcribed into an access database. The process followed for the chart review is explained in further detail in Rozich et al [106].

Table 3.2 Harm Categories based on the NCC MERP classification of errors

Category	Description
A	Circumstances or events that have the capacity to cause error
B	An error occurred but the error did not reach the patient (An "error of omission" does reach the patient)
C	An error occurred that reached the patient but did not cause patient harm
D	An error occurred that reached the patient and required monitoring to confirm that it resulted in no harm to the patient and/or required intervention to preclude harm
E	An error occurred that may have contributed to or resulted in temporary harm to the patient and required intervention
F	An error occurred that may have contributed to or resulted in temporary harm to the patient and required initial or prolonged hospitalization
G	An error occurred that may have contributed to or resulted in permanent patient harm
H	An error occurred that required intervention necessary to sustain life
I	An error occurred that may have contributed to or resulted in the patient's death

Data Description

The final dataset provided by the CHOA researchers did not include any patient identifiers (a HIPAA waiver was approved in the GT IRB protocol). A total of 2278

triggers were identified (including T0: No triggers found). The data included the following variables:

- Patient Number: It associates the trigger to an individual patient
- Time period: One of the six time periods of the data collection
- Patient group: There were 4 groups, a combination of the two areas (General or Intensive Care) and two hospitals (Egleston or Scottish Rite)
- Type of event: According to the trigger review, there were three options: 'No event', 'ADE' or 'Error'
- Harm category: From A to I
- Doses per day (for a 14-day period)
- Date of birth
- Date of admission
- Date of discharge

3.2.2 Data Analysis

The following additional variables were calculated from the collected data:

- Age: This is the patient age at time of admission
- Length of Stay: Number of days between date of admission and discharge
- Reviewed Days: This is equal to Length of Stay if less than 14, otherwise it is 14
- Number of errors and ADEs for each harm category (calculated by patient)
- Total number of doses: The total number of doses a patient was given during the reviewed days. This variable was used as the *exposure variable,* based on the assumption that the probability of an error or ADE is the same for each dose

- Rate of med errors and ADEs: That is the number of medication errors or ADEs per dose.

Length of stay and age were transformed into categorical variables. Length of stay was grouped into two categories:

- L1: 1 to 7 days
- L2: 8 or more days

The following pediatric age group classification was used:

- Neonates: 0 to 27 days
- Infants and toddlers: 28 days to 23 months
- Children: 2 to 11 years
- Adolescents: 12 to 18 years

Outcome measures

In addition to modeling the number of all ADES and medication errors, the pharmacists at CHOA were also interested in examining the events that reached the patient and had potential for harm, i.e., categories D and above. Therefore, the following outcome measures were examined:

1. Rate of all medication errors
2. Rate of all ADEs
3. Rate of medication errors category D or higher
4. Rate of ADEs category D or higher

For each time period and patient subgroup (by hospital and by care area), the rates of medication errors and ADEs per 1000 doses were calculated and trends across time periods examined.

Potential initial predictors

Based on the literature review, potential associations between each outcome measure and the following variables were assessed:

1. Time period: Six times periods, from T0 to T6
2. Hospital: Two hospitals (EG or SR)
3. Care area: General care (GCA) or Intensive care (ICU)
4. Age: The age of the patient at admission

Data models

The following nonlinear models were constructed for each outcome measure:

1. Poisson model (P)
2. Negative Binomial (NB)
3. Hurdle Poisson (HP)
4. Hurdle Negative Binomial (HNB)
5. Zero-inflated Poisson (ZIP)
6. Zero-inflated Negative Binomial (ZINB)

Nested models (e.g., Poisson vs. Negative Binomial) were compared using the Likelihood ratio test. Nonnested models (e.g., Poisson vs. Zero-inflated Poisson) were compared using the Vuong test. All models were compared using the Akaike Information Criteria (AIC). The statistical modeling was performed in version 2.9.2 of R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna, Austria).

3.3 Results and Discussion

A total of 959 patients were randomized and evaluated for possible ADEs and medication errors using the trigger tool. Approximately 160 patient charts were examined for each time period (except at baseline): forty patients each from GCA and ICU from both campuses were reviewed. The final dataset had 937 inpatient charts. Charts were excluded for the following reasons:

- 10 charts had missing data: medication data missing or incomplete (7), labs data missing (1), date of birth missing (1), time period missing (1)
- 8 charts had 0 dose count
- 3 charts had length of stay < 2 days
- 1 chart was a duplicate

Of the 937 examined patient charts, 80 (8.5%) presented at least one medication error, and 223 (23.8%) had at least one adverse drug event (Table 3.3). The mean number of medication errors was 0.1 and of ADEs was 0.34.

Table 3.3 Frequency distribution of medication errors and ADEs

Number of errors	Frequency	Percent
0	857	91.5%
1	70	7.5%
2	7	0.7%
3	2	0.2%
4	1	0.1%

Number of ADEs	Frequency	Percent
0	714	76.2%
1	154	16.4%
2	43	4.6%
3	22	2.3%
4	4	0.4%

Table 3.4 presents the descriptive statistics for the independent variables. The distribution of patients and the percentage of patients that experienced one or more incidents are presented for each category of the variables.

Table 3.4 Descriptive statistics for independent variables

Variable		Number of patients (%)	% of patients with 1 or more med errors	% of patients with 1 or more ADEs
Age	Neonates	129 (13.8)	5.4	21.7
	Infants & toddlers	280 (29.9)	7.1	20.4
	Children	333 (35.5)	9.6	22.8
	Adolescents	171 (18.2)	9.4	29.2
	Adults	24 (2.6)	20.8	50.0
Time	T0	88 (9.4)	17.0	12.5
	T1	174 (18.6)	9.2	14.4
	T2	160 (17.1)	10.6	28.8
	T3	163 (17.4)	14.1	30.7
	T4	177 (18.9)	3.4	29.4
	T5	175 (18.7)	1.7	22.3
Hospital	EG	459 (49.0)	7.2	29.2
	SR	478 (51.0)	9.8	18.6
Area	GCA	440 (47.0)	5.9	19.8
	ICU	497 (53.0)	10.9	27.4
LOS2	1 to 7 days	653 (69.7)	6.9	15.8
	More than 7 days	284 (30.3)	12.3	42.3

The average length of stay was 11.6 days, ranging from 2 to 263 days. Most of the patients in the study were hospitalized less than a week. The overall mean length of stay did not vary significantly over time, except during T4 and T5. The community-based hospital (SR) had equal or higher overall LOS than the academic-based one, except during T5. The LOS in the ICUs was always higher than in the GCA, and that difference was maintained at every time period.

Table 3.5 Average length of stay in days by hospital and area over time

	T0	T1	T2	T3	T4	T5	Total
EG	6.8	8.4	7.2	9.4	13.6	19.6	11.4
GCA	3.3	2.7	4.6	4.8	5.6	6.4	4.7
ICU	10.6	13.8	9.6	13.8	20.0	31.5	17.6
SR	9.7	10.4	6.8	10.2	17.2	15.0	11.8
GCA	3.6	3.0	2.8	3.7	6.0	4.1	3.9
ICU	13.0	18.0	10.6	15.4	27.1	23.7	18.5
Total	8.4	9.4	7.0	9.8	15.4	17.4	11.6

About four fifths of ADEs (78%) belonged to harm categories D and above, while only two fifths of medication errors (39%) belonged to the same categories

The rates of medication errors and ADEs seen from Figure 3.3 to Figure 3.6 were calculated using the values in Table 3.6.

Table 3.6 Distribution of medication errors and ADEs by patient subgroup

Time Period	Hospital	Area	Charts	Total Doses	All Errors	Errors D+	All ADEs	ADEs D+
Total			937	86323	94	37	322	251
T0	EG	GCA	20	603	1	1	7	7
		ICU	19	1744	4	2	2	2
	SR	GCA	17	458	4	0	5	5
		ICU	32	3020	8	1	4	3
T1	EG	GCA	41	1182	2	0	9	9
		ICU	43	4563	2	2	14	14
	SR	GCA	46	905	2	0	3	2
		ICU	44	6214	12	1	10	9
T2	EG	GCA	39	2351	6	1	14	7
		ICU	40	4245	4	3	15	6
	SR	GCA	40	1288	4	0	10	8
		ICU	41	3874	5	1	24	18
T3	EG	GCA	40	3431	10	6	23	11
		ICU	42	5771	6	2	36	26
	SR	GCA	36	924	0	0	12	7
		ICU	45	6326	15	12	19	17
T4	EG	GCA	38	3937	2	1	17	14
		ICU	48	10022	1	1	25	24
	SR	GCA	43	1534	1	0	7	6
		ICU	48	7273	2	0	15	13
T5	EG	GCA	42	1790	0	0	10	10
		ICU	47	8016	1	1	22	15
	SR	GCA	38	1661	0	0	7	6
		ICU	48	5191	2	2	12	12

The rate of medication errors showed an overall potential trend downwards across time (Figure 3.3). Overall error rates dropped to almost zero at T4 and T5. The error rate

for high severity categories (D and above) was initially quite low and remained low throughout the study period, except for an increase in the error rates of the community-based hospital during T3. The first reduction in the rate of medication errors occurred after the implementation of EpicRx, and the second reduction was after implementing the physician documentation function. Initial differences between hospitals disappeared over time.

Surprisingly, the error rates for the ICUs were lower than for GCAs, except during T3, when the rate for the GCA in SR was 0. The largest reduction in medication error rates over time was observed in the general care areas for both hospitals, in particular for the community-based one (Figure 3.4).

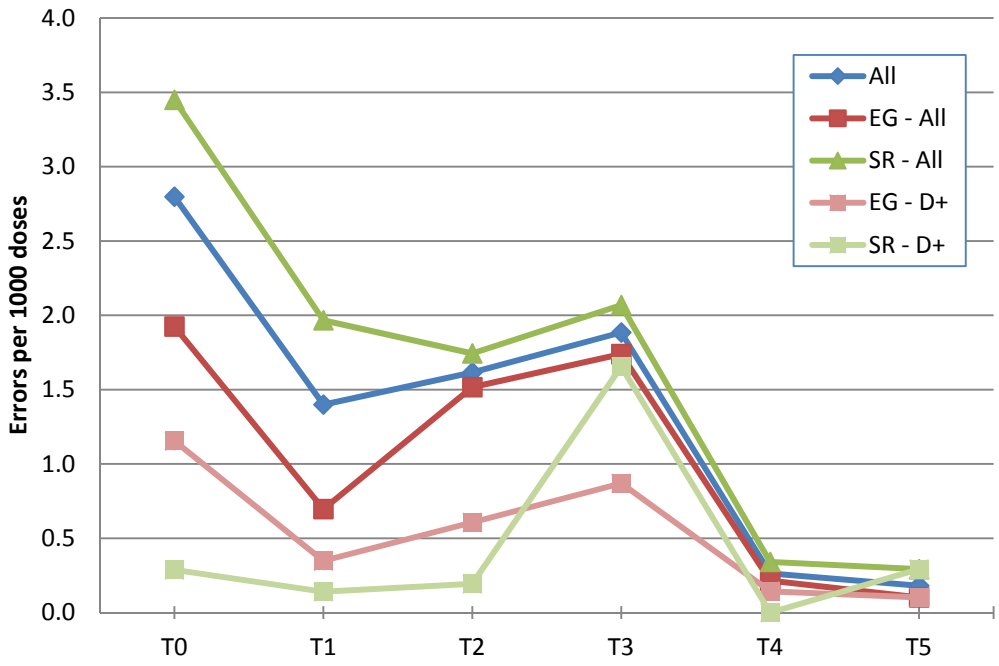

Figure 3.3 Medication error rates by hospital type

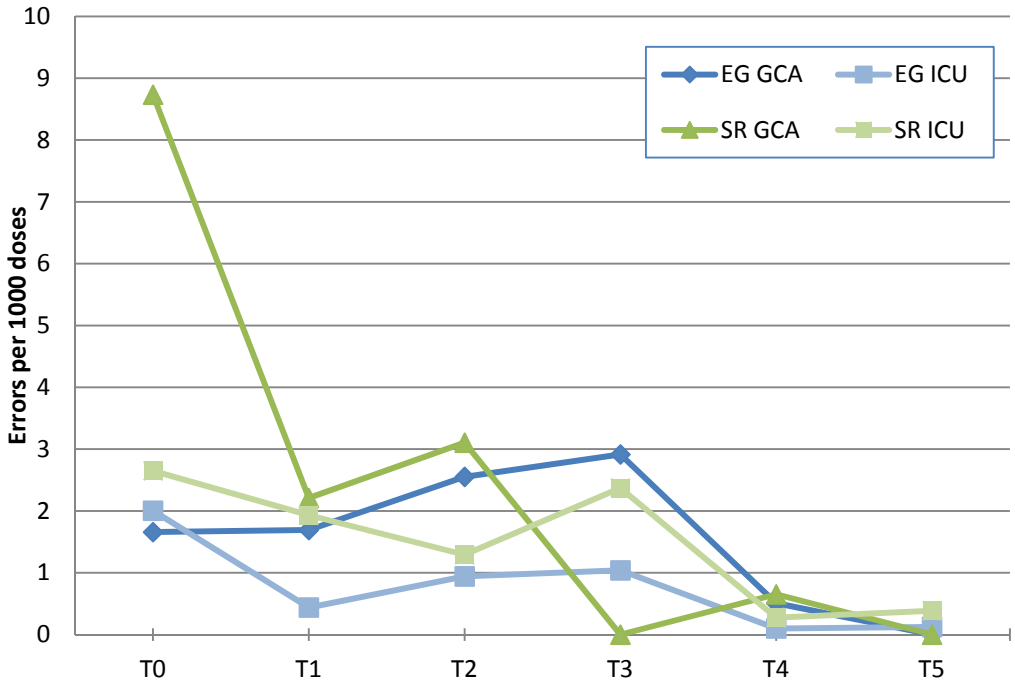

Figure 3.4 Medication error rates by hospital and care area

The rate of ADEs was higher than the rate of medication errors. ADE rates, since they include known, expected medication side effects, were more variable across the study period. The most variability occurred during T2 and T3. The rate of overall ADEs did not decrease at the end of the EMR implementation with respect to baseline, and it actually increased during T2 and T3. Initial differences in the rate of ADEs between hospitals disappeared by T5 (Figure 3.5).

At the beginning of the study, the rate of ADEs in the GCAs was much higher than in the ICUs. The rate of ADEs in the GCAs decreased drastically from T0 to T1. After T1, the rate of ADEs either remained the same or increased, except for the GCA at EG, where we observed a slight decreasing trend in the rate of ADEs. The initial differences between care units decreased by T4 and ADE rates remained almost the same during T5 (Figure 3.6).

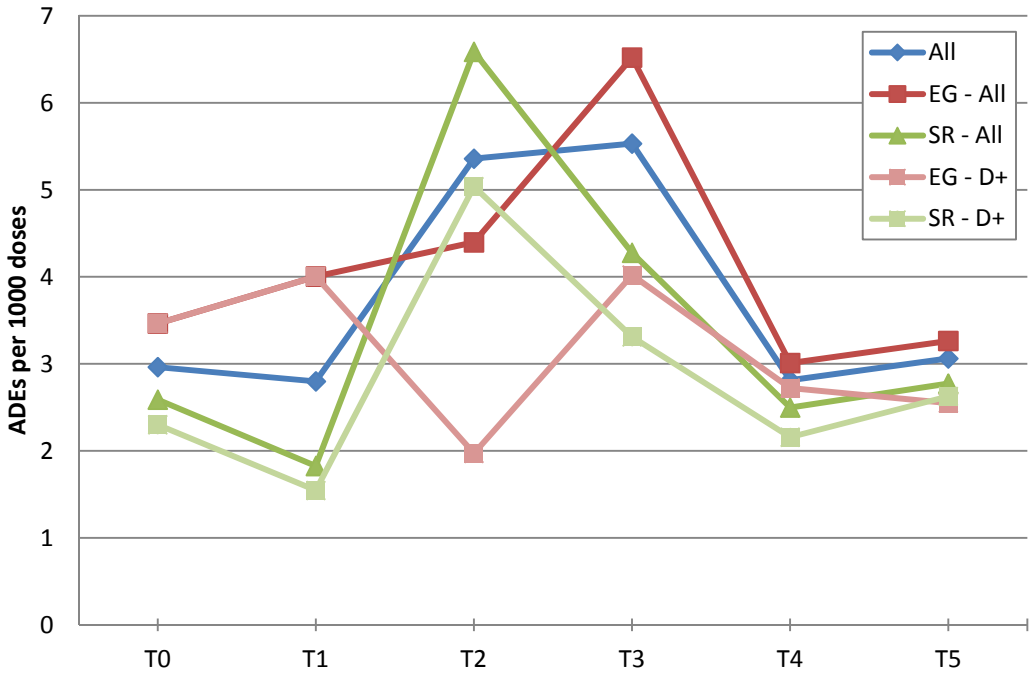

Figure 3.5 Rate of ADEs (All and D+) by hospital type

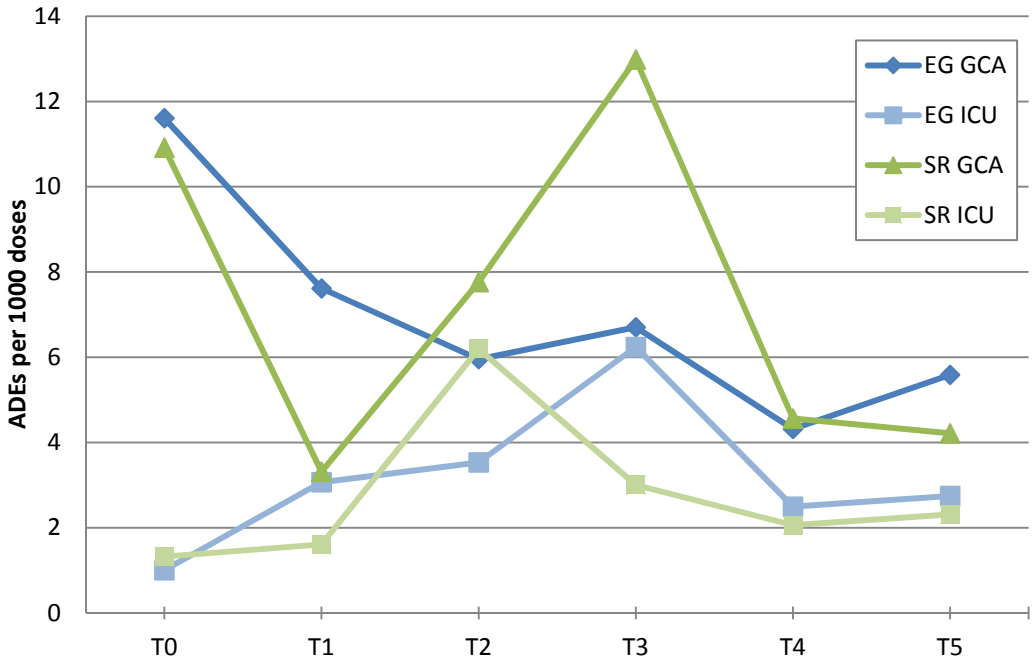

Figure 3.6 Rate of ADEs by hospital and care area

During the implementation of the EMR, CHOA initiated other projects in addition to EMR to reduce ADEs and medication errors. During August 2005 (between T1 and T2), smart pump technology for syringes was implemented to prevent medication errors at the bedside by catching any miskeys and/or pump programming errors. Also, in summer 2006 (between T2 and T3), the pharmacy completed a Six Sigma project targeting a reduction in medication errors in preparation and dispensing medications. The phased EMR implementation and these other initiatives contributed to the changes in medication error rates and ADEs observed in this study.

3.3.1 Models Estimation and Evaluation

The distribution of both the number of medication errors and adverse drug events followed the expected count distribution with a large number of zeros, which then drastically drops for non zero events, in particular for medication errors.

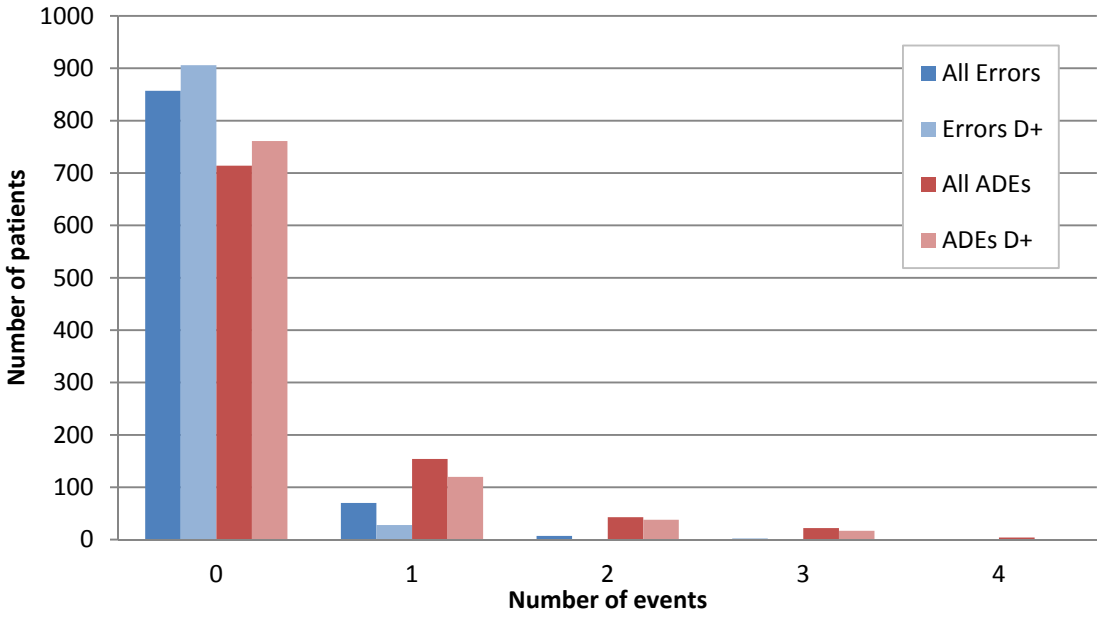

Figure 3.7 Distribution of total medication errors and adverse drug events

Six different nonlinear regression models for this type of data were evaluated: Poisson, Negative Binomial, ZI Poisson, ZI Negative Binomial, Hurdle Poisson and Hurdle Negative Binomial. Each model is applied to the four outcomes measures: overall errors, overall ADEs, high severity errors and high severity ADEs (category D and above). The variable age category was not included in the set of predictors because it was not significant in an initial univariate analysis. The three variables included in all the models as candidate predictors were: Time, Hospital and Area. A description of each variable is presented in Table 3.7.

Table 3.7 Independent Variables

Variable	Description
T1*	equals 1 if the patient stayed in the hospital during period T1, 0 otherwise
T2*	equals 1 if the patient stayed in the hospital during period T2, 0 otherwise
T3*	equals 1 if the patient stayed in the hospital during period T3, 0 otherwise
T4*	equals 1 if the patient stayed in the hospital during period T4, 0 otherwise
T5*	equals 1 if the patient stayed in the hospital during period T5, 0 otherwise
Hospital	equals 1 if the patient stayed at Scottish Rite, 0 if he stayed at Egleston
Area	equals 1 if the patient stayed in the ICU, 0 if he stayed in the GCA

**The time variables T1-T5 are dummy variables, with T0 as the reference value.*

The time categories represent the patients that were hospitalized during that specific period of time. Table 3.8 presents the parts of the EMR system that were live at each time period. The 'x' indicates that the system functionality was live and the '--' indicates that the functionality had not been implemented yet.

Table 3.8 Description of the systems implemented at each collection time

System functionality	Time Period					
	T0	T1	T2	T3	T4	T5
Epic Rx	--	x	x	x	x	x
Smart pumps, eMAR and clinical ancillary orders	--	--	x	x	x	x
Patient safety initiatives	--	--	--	x	x	x
RN and MD doc and non-clinical ancillary orders	--	--	--	--	x	x
CPOE	--	--	--	--	--	x

Medication Errors

Table 3.9 presents the parameters estimates and their standard errors for the six regression models of all medication errors. An initial evaluation of the models was performed by comparing their AIC values (lower is the best), and the following models were selected: Poisson, NB, ZIP and ZINB. Further examination using the likelihood ratio test to compare nested models (P vs. NB, ZIP vs. ZINB, and P vs. HP) gave significant results for P vs. NB, indicating overdispersion. The Vuong test, used to compare non-nested models, was not significant for any of the comparisons: P vs. ZIP, NB vs. ZINB and NB vs. HNB. This indicated that a two-part model was not necessary to model the high number of zeros in the data. From these tests, we eliminated the zero-inflated models.

The next step in the selection of the best model involves a comparison of the fitted and observed values. We observed in Table 3.10 that the Poisson model predicted better the number of patients with no medication errors and those with one error. The NB model overestimated the number of zeros and underestimated the number of patients with one medication error. Therefore, the Poisson model was selected as the best fit to model the number of all medication errors.

Table 3.9 Parameter estimates and standard errors for all six count models

Predictor		Poisson	NB	ZIP	ZINB	HP	HNB
Poisson or Negative Binomial							
(Intercept)	Coef	**-5.68**	**-5.57**	**-5.07**	**-5.42**	**-5.28**	-16.45
	S.E.	**0.31**	**0.35**	**0.43**	**0.54**	**0.76**	171.24
T1	Coef	**-0.70**	**-0.78**	**-0.82**	**-0.76**	-0.55	-0.47
	S.E.	**0.34**	**0.39**	**0.38**	**0.39**	0.98	1.22
T2	Coef	-0.61	-0.66	-0.69	-0.65	-0.64	-1.43
	S.E.	0.34	0.38	0.37	0.40	0.98	1.41
T3	Coef	-0.41	-0.42	-0.46	-0.39	0.05	0.70
	S.E.	0.30	0.35	0.35	0.37	0.76	1.02
T4	Coef	**-2.34**	**-2.46**	**-2.46**	**-2.45**	-12.52	-14.44
	S.E.	**0.48**	**0.50**	**0.50**	**0.52**	343.23	1430.65
T5	Coef	**-2.73**	**-2.85**	**-2.82**	**-2.80**	-9.47	-6.78
	S.E.	**0.63**	**0.65**	**0.65**	**0.66**	131.68	33.71
HospitalSR	Coef	**0.52**	**0.52**	0.71	**0.62**		
	S.E.	**0.22**	**0.24**	0.53	**0.29**		
Area ICU	Coef	**-0.63**	**-0.60**	-0.87	-0.84	-0.83	-1.51
	S.E.	**0.22**	**0.25**	0.57	0.53	0.54	0.85
Log(theta)					-0.04		-11.89
					0.58		171.24
Logistic model (Zero inflation part)							
(Intercept)	Coef			-0.43	-2.42	**-2.02**	**-2.02**
	S.E.			0.84	5.22	**0.33**	**0.33**
T1	Coef					-0.66	-0.66
	S.E.					0.39	0.39
T2	Coef					-0.51	-0.51
	S.E.					0.39	0.39
T3	Coef					-0.20	-0.20
	S.E.					0.36	0.36
T4	Coef					**-1.76**	**-1.76**
	S.E.					**0.50**	**0.50**
T5	Coef					**-2.46**	**-2.46**
	S.E.					**0.65**	**0.65**
HospitalSR	Coef			0.49	1.58		
	S.E.			1.52	4.36		
Area ICU	Coef			-0.74	-9.57	**0.69**	**0.69**
	S.E.			1.43	169.82	**0.25**	**0.25**
AIC		591.97	584.78	591.30	590.24	613.89	607.43

Values in bold are significant for p<= 0.05

Hospital type, care area and time periods T1, T4 and T5 were significant in the Poisson model (Table 3.9). Patients in the community-based hospital (SR) had a higher rate of errors compared to the academic hospital, and patients in the ICUs had a lower error rate than those in the GCAs. The latter result might be due to a higher nurse-patient ratio in those units as well as the usually higher number of doses per patient in the ICU (Table 3.6). The effect of time period was significant for T1 (post EpicRX), T4 (post Phys documentation) and T5 (post CPOE). The rate of errors at T1 was reduced by half compared to the baseline. During T4 and T5, the errors rate decreased to 10% and 7% of the baseline rate, respectively. Reviewing the patient charts, we observed that most of the medication errors found at baseline were "order entry errors", which were completely eliminated with the implementation of CPOE. These results confirmed previous studies that indicated CPOE as the tool that help decrease medical errors the most [78, 107-109].

Table 3.10 Observed and predicted number of medication errors

Number of errors	No. of patients observed	No. of patients predicted					
		P	NB	ZIP	ZINB	HP	HNB
0	857	858	862	861	862	857	857
1	70	67	58	60	58	74	73
2	7	9	11	12	11	5	5
3	2	2	3	3	3	1	1
4	1	0	1	1	1	0	0
5	0	0	1	0	0	0	0

The results from the model estimates can be confirmed in Figure 3.8, which displays the boxplots of the rate of errors (errors per 1000 doses) for each predictor.

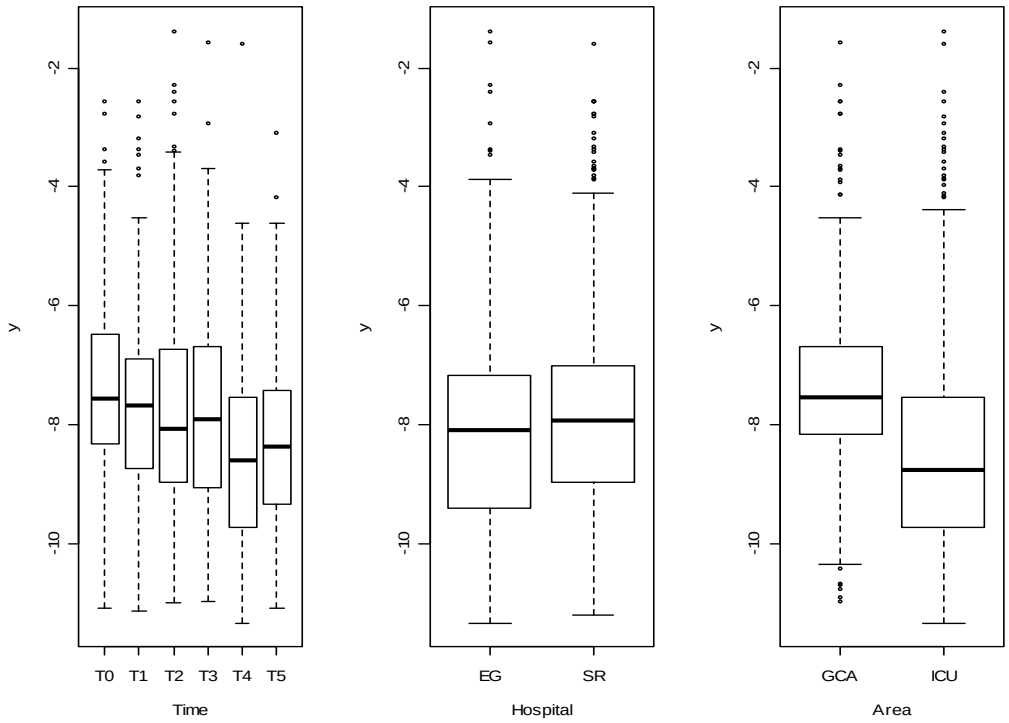

Figure 3.8 Box plots of error rates vs. independent variables
y = log(errors+0.01)-log(doses)

High severity medication errors (D and above)

Table 3.11 presents the results of the six models for medication errors category D and above. Based on the AIC values, the NB and the ZIP are the best-fitting models. The ZINB could not be fitted to the data. This happened because for some of the predictors, there were so few non-zero values that the parameters of the ZINB could not be estimated. For instance, for the Time variables, except for T3, all the rest of them had zero values for frequencies higher than 1, which implied a zero variance for non-zero frequencies.

Table 3.11 Parameter estimates and standard errors for errors D and above

Predictor		Poisson	NB	ZIP	ZINB	HP	HNB
Poisson or Negative Binomial							
(Intercept)	Coef	**-7.30**	**-6.99**	**-4.92**	NA	-12.70	-19.98
	S.E.	**0.61**	**0.71**	**0.76**		59.26	39.97
T1	Coef	-1.08	-1.20	-0.45	NA	-2.50	-2.25
	S.E.	0.76	0.87	1.00		323.55	596.08
T2	Coef	-0.48	-0.46	0.21	NA	-2.46	-3.49
	S.E.	0.67	0.77	0.92		424.06	293.25
T3	Coef	0.57	0.76	-0.21	NA	7.70	9.70
	S.E.	0.55	0.67	0.77		59.26	40.00
T4	Coef	**-2.06**	**-2.26**	-1.98	NA	-3.07	-4.87
	S.E.	**0.87**	**0.97**	1.26		272.20	382.14
T5	Coef	-1.34	-1.49	-1.46	NA	-2.50	-2.51
	S.E.	0.76	0.87	1.21		453.78	966.53
HospitalSR	Coef	-0.01	-0.42	**1.97**	NA	11.18	11.96
	S.E.	0.34	0.42	**0.65**		323.46	300.96
Area ICU	Coef	0.03	0.05	**-2.46**	NA	-11.99	-15.36
	S.E.	0.39	0.46	**0.67**		323.46	300.96
Log(theta)							NA
Logistic model (Zero inflation part)							
(Intercept)	Coef			**1.97**	NA	**-3.85**	**-3.85**
	S.E.			**1.15**		**0.67**	**0.67**
T1	Coef			1.14	NA	-0.94	-0.94
	S.E.			1.57		0.78	0.78
T2	Coef			0.80	NA	-0.33	-0.33
	S.E.			1.32		0.69	0.69
T3	Coef			-1.46	NA	0.72	0.72
	S.E.			1.15		0.59	0.59
T4	Coef			0.15	NA	-1.42	-1.42
	S.E.			1.95		0.88	0.88
T5	Coef			-0.54	NA	-1.00	-1.00
	S.E.			1.82		0.78	0.78
HospitalSR	Coef			13.15	NA	-0.29	-0.29
	S.E.			143.95		0.37	0.37
Area ICU	Coef			-13.62	NA	**1.37**	**1.37**
	S.E.			143.95		**0.46**	**0.46**
AIC		303.56	289.77	298.05	NA	315.58	302.22

Results in bold are significant for p<=0.05

The likelihood ratio test comparing the P to the NB model was significant ($p<0.001$), indicating overdispersion. The Vuong test comparing P vs. ZIP was also significant, indicating that the overdispersion was due to an unusual high number of zeros. Table 3.12 displays the predicted values of each model, and we observed that the ZIP predicts better than the NB the non-zero error frequencies. Then, we selected the ZIP as the model that best fit the number of medication errors of category D and above.

There were no significant factors in the logit part of the ZIP model. Type of hospital and area were both significant in the Poisson part of the model. Patients in the community-based hospital had a higher rate of errors category D or above, and patients in the ICU had a lower rate than those in general care. The results from the model estimates can be confirmed in Figure 3.9, which displays the boxplots of the rate of errors for each independent predictor. The selection of a ZI model with non significant predictors in the logistic part implies that the sample population is heterogeneous, and that there is an unobserved group of patients that never experience high category medication errors ("fixed" zeros) who we are not able to describe with the current predictors in our model.

Table 3.12 Observed and predicted number of high severity errors

Number of errors D+	No. of patients observed	No. of patients predicted					
		P	NB	ZIP	ZINB	HP	HNB
0	906	903	907	907	NA	906	906
1	28	31	22	23	NA	28	28
2	1	3	4	4	NA	1	1
3	1	0	2	1	NA	0	0
4	1	0	1	1	NA	0	0
5	0	0	0	0	NA	0	0

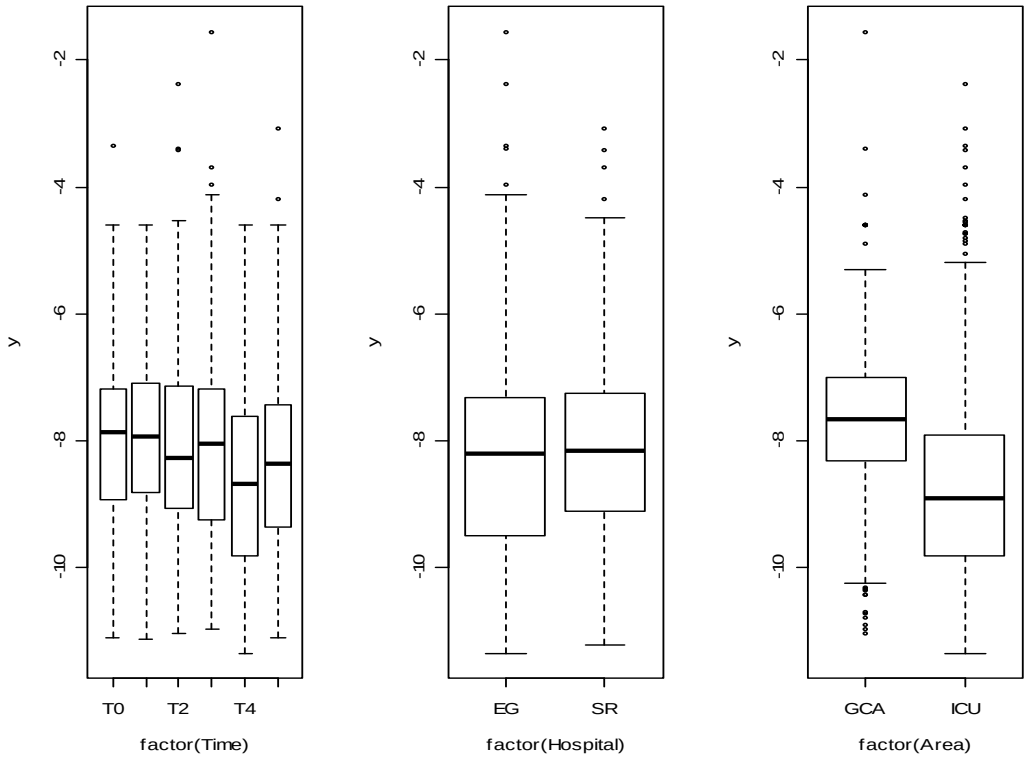

Figure 3.9 Box plots of high severity errors rates vs. each independent predictor
y = log(errors_Dplus + 0.01) - log(doses)

Adverse Drug Events

Table 3.13 presents the results of the six models for all categories of adverse drug events. The models with the best fit, based on their AIC value, were the NB, ZIP, ZINB followed by the Poisson model. The result of the test comparing the Poisson to the negative binomial model was significant, indicating overdispersion of the observed data. The tests comparing the simple models to their zero-inflated counterparts were also significant, implying that a ZI model would fit the data better.

Table 3.13 Parameter estimates and standard errors for all ADEs

Predictor		Poisson	NB	ZIP	ZINB	HP	HNB
Poisson or Negative Binomial							
(Intercept)	Coef	**-5.21**	**-5.06**	**-4.28**	**-4.32**	**-4.32**	**-5.66**
	S.E.	**0.25**	**0.29**	**0.35**	**0.43**	**0.37**	**1.72**
T1	Coef	-0.08	-0.02	-0.29	-0.22	-0.31	0.02
	S.E.	0.28	0.32	0.42	0.52	0.45	0.79
T2	Coef	0.43	0.50	**-0.70**	**-0.48**	**-0.90**	-0.77
	S.E.	0.27	0.30	**0.36**	**0.45**	**0.41**	0.72
T3	Coef	**0.48***	**0.59**	-0.39	-0.19	-0.48	0.05
	S.E.	**0.27**	**0.30**	0.36	0.45	0.37	0.71
T4	Coef	-0.16	-0.14	**-1.17**	**-1.10**	**-1.65**	**-1.79**
	S.E.	0.27	0.30	**0.38**	**0.45**	**0.44**	**0.73**
T5	Coef	-0.05	-0.04	**-1.04**	**-0.97**	**-1.21**	-1.17
	S.E.	0.27	0.31	**0.39**	**0.16**	**0.44**	0.76
HospitalSR	Coef	-0.11	-0.20	0.22	-0.06	**0.40**	0.28
	S.E.	0.12	0.13	0.15	0.17	**0.20**	0.34
Area ICU	Coef	**-0.66**	**-0.74**	**-0.53**	**-0.61**	**-0.55**	**-0.72**
	S.E.	**0.12**	**0.14**	**0.14**	**0.16**	**0.20**	**0.34**
Log(theta)					**1.02**		-1.68
					0.47		2.02
Logistic model (Zero inflation part)							
(Intercept)	Coef			-0.53	-0.69	**-1.91**	**-1.91**
	S.E.			0.78	1.04	**0.34**	**0.34**
T1	Coef			-0.46	-0.40	0.17	0.16
	S.E.			0.74	0.90	0.39	0.39
T2	Coef			-13.11	-16.25	**1.05**	**1.05**
	S.E.			132.35	696.39	**0.36**	**0.37**
T3	Coef			**-1.88**	**-1.90**	**1.13**	**1.13**
	S.E.			**0.71**	**0.91**	**0.37**	**0.37**
T4	Coef			**-2.86**	-10.68	**1.07**	**1.07**
	S.E.			**1.22**	207.28	**0.37**	**0.36**
T5	Coef			**-2.46**	-3.71	**0.68**	0.68
	S.E.			**1.25**	4.95	**0.37**	0.37
HospitalSR	Coef			**1.33**	1.14	**-0.60**	**-0.60**
	S.E.			**0.51**	0.66	**0.16**	**0.16**
Area ICU	Coef			0.69	0.69	**0.45**	**0.45**
	S.E.			0.63	0.80	**0.16**	**0.16**
AIC		1372.20	1344.00	1347.84	1342.91	1427.87	1411.52

Results in bold are significant for p <= 0.05
*The p value for the coefficient of T3 was 0.06

However, when looking at how well the models predicted the number of ADEs, we observed that the NB and ZI models overestimated the number of zeros and underestimated the number of patients with 1 ADE (Table 3.14). The Poisson model estimated more accurately the zero and the non-zero counts. Therefore, we selected the Poisson model as the best fit for modeling the number of ADEs.

Table 3.14 Observed and predicted number of adverse drug events

Number of ADEs	No. of patients observed	No. of patients predicted					
		P	NB	ZIP	ZINB	HP	HNB
0	714	718	729	730	731	714	714
1	154	155	135	136	132	183	181
2	43	41	40	43	41	28	27
3	22	13	16	16	16	7	8
4	4	5	7	7	7	2	3
5	0	2	4	0	0	0	0
6	0	1	2	0	0	0	0
7	0	1	1	0	0	0	0
8	0	0	1	0	0	0	0
9	0	0	1	0	0	0	0

Only T3 and type of care area were significant in the Poisson model. The significant coefficients indicate the rate of ADEs was higher during T3 compared to baseline, and it was lower for patients in the ICU compared to those in the GCA. We can verify that the model reflects the behavior of the observed data by examining the plots in Figure 3.10.

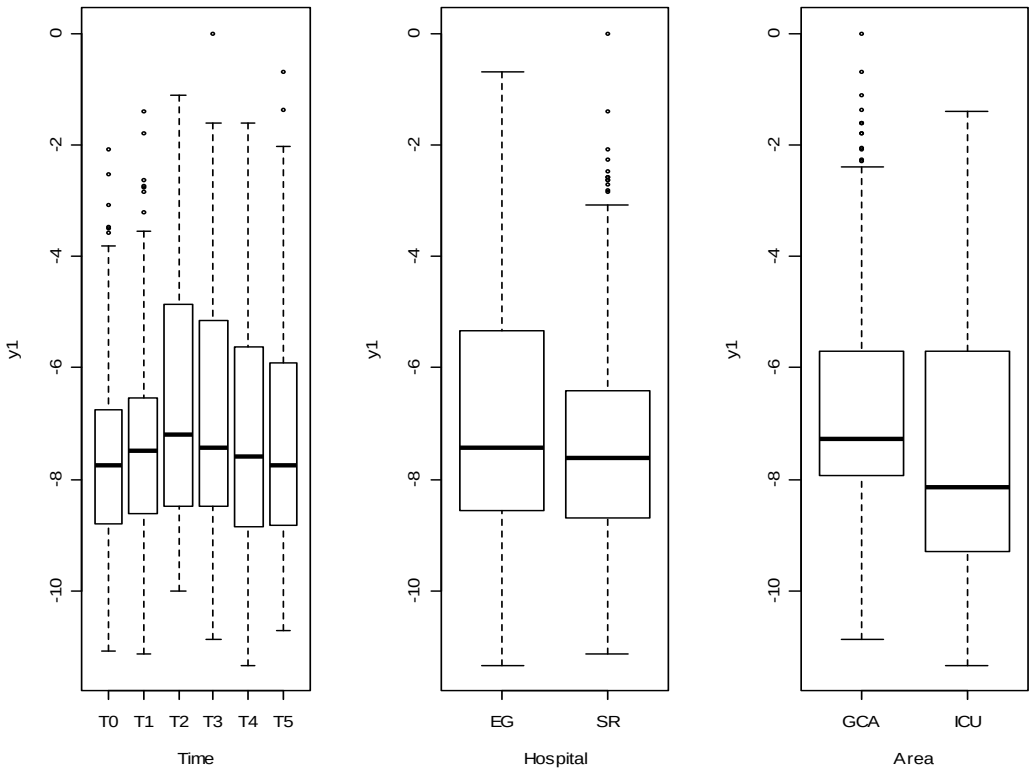

Figure 3.10 Bivariate plots of rate of ADEs vs. predictors
y1=log(ADEs+0.01)-log(Doses)

Since HITs are supposed to decrease medication errors and ADEs [65, 110], we decided to further examine the data to understand what happened during T3. Looking at the patient charts, we observed a peak in the number of ADEs identified in the Lab Diagnostics, Medications and Nursing ancillary flowsheet sections of the charts (Table 3.15). We could credit this increase to a better documentation of medications after the implementation of the eMAR and ancillary documentation functionalities of the EMR. In addition, patient safety initiatives were in place from T3 onwards, which may have also encouraged a more accurate reporting and documentation.

Table 3.15 Distribution of ADEs identified in the different chart sections

Chart Section	T0	T1	T2	T3	T4	T5
Discharge Summary	1	6	4	8	17	14
H & P	0	0	1	0	1	0
Lab Diagnostics	2	2	**8**	**17**	3	2
Medications	8	16	**26**	**31**	28	20
Nursing Ancillary Flowsheet	5	8	**17**	**23**	0	2
Operative Record	1	0	3	0	0	0
Orders	1	4	4	12	15	13
Total	**18**	**36**	**63**	**91**	**64**	**51**

High Severity ADEs

Table 3.16 presents the results of the six models for adverse drug events category D and above. Following the previous approach to select a model, the models with the best fit (lower AIC) are the NB, ZIP and ZINB. The likelihood tests comparing P vs. NB and ZIP vs. ZINB were both significant. Likewise, the Vuong test for P vs. ZIP and NB vs. ZINB were also significant, indicating that the overdispersion might be due to an excess of zeros.

Table 3.16 Parameter estimates and standard errors for ADEs D+

Predictor		Poisson	NB	ZIP	ZINB	HP	HNB
Poisson or Negative Binomial							
(Intercept)	Coef	**-5.33**	**-5.22**	**-4.36**	**-4.58**	**-4.27**	**-5.39**
	S.E.	**0.27**	**0.03**	**0.34**	**0.49**	**0.37**	**1.75**
T1	Coef	-0.08	0.02	-0.24	-0.14	-0.34	-0.04
	S.E.	0.30	0.35	0.42	0.57	0.45	0.82
T2	Coef	0.03	0.12	**-1.07**	-0.84	**-0.87**	-0.68
	S.E.	0.29	0.35	**0.41**	0.53	**0.45**	0.79
T3	Coef	0.17	0.23	**-0.79**	-0.55	**-0.77**	-0.40
	S.E.	0.27	0.34	**0.37**	0.52	**0.39**	0.76
T4	Coef	-0.20	-0.15	**-1.18**	**-0.99**	**-1.74**	**-2.02**
	S.E.	0.28	0.33	**0.38**	**0.53**	**0.46**	**0.78**
T5	Coef	-0.14	-0.15	**-1.15**	**-0.97**	**-1.16**	-1.04
	S.E.	0.28	0.33	**0.40**	**0.53**	**0.45**	0.79
HospitalSR	Coef	-0.03	-0.80	**0.50**	0.28	0.34	0.09
	S.E.	0.13	0.16	**0.19**	0.23	0.23	0.40
Area ICU	Coef	**-0.62**	**-0.66**	**-0.47**	**-0.50**	**-0.50**	-0.67
	S.E.	**0.13**	**0.16**	**0.19**	**0.20**	**0.23**	0.39
Log(theta)					0.36		-1.56
					0.45		2.16
Logistic model (Zero inflation part)							
(Intercept)	Coef			-0.06	-1.13	**-2.16**	**-2.17**
	S.E.			0.70	1.45	**0.36**	**0.37**
T1	Coef			-0.35	-0.32	0.19	0.19
	S.E.			0.66	1.02	0.41	0.41
T2	Coef			**-2.59**	-15.41	0.52	0.52
	S.E.			**1.39**	995.00	0.39	0.40
T3	Coef			**-1.78**	-2.23	**0.79**	**0.80**
	S.E.			**0.73**	1.26	**0.38**	**0.39**
T4	Coef			**-2.17**	-2.00	**1.04**	**1.05**
	S.E.			**0.78**	1.97	**0.38**	**0.38**
T5	Coef			**-1.90**	-2.94	0.55	0.55
	S.E.			**0.94**	2.54	0.39	0.39
HospitalSR	Coef			**1.33**	**1.73**	**-0.39**	**-0.39**
	S.E.			**0.58**	**0.92**	**0.17**	**0.17**
Area ICU	Coef			0.35	0.71	**0.52**	**0.52**
	S.E.			0.57	1.00	**0.17**	**0.17**
AIC		1203.80	1162.60	1169.57	1163.35	1233.22	1222.46

Results in bold are significant for p<=0.05

When looking at how well the models predicted the actual frequency of ADEs, the ZIP estimated more accurately the number of zeros. Similarly, although it underestimated the distribution of the non-zeros, the ZIP was the one closer to the actual counts (Table 3.17). Therefore, we selected the ZIP model as the best fit to predict the frequency of ADEs from categories D and above. The selection of the ZIP model implies that the observed data is heterogeneous, with a subgroup of patients that will never experience those types of ADEs.

Table 3.17 Observed and predicted number of D+ ADEs

Number of ADEs D+	No. of patients observed	No. of patients predicted					
		P	NB	ZIP	ZINB	HP	HNB
0	761	752	769	767	769	761	761
1	120	140	111	114	111	145	143
2	38	32	31	35	33	22	21
3	17	9	12	13	12	6	6
4	1	3	6	5	6	2	3
5	0	1	3	0	0	0	0
6	0	0	2	0	0	0	0
7	0	0	1	0	0	0	0
8	0	0	1	0	0	0	0

In the logistic component of the ZIP model, hospital type and time periods T2, T3, T4 and T5 were significant (see Table 3.16). The positive coefficient for HospitalSR means that patients from Scottish Rite (community-based) had a higher probability of belonging to the group that never experiences high severity ADEs. On the contrary, the negative coefficients for T2, T3, T4 and T5 indicate that patients hospitalized during those time periods have a higher probability of having an ADEs from category D or above compared to those at baseline. In the Poisson component, all the predictors were significant except for T1. The sign of the coefficients indicates that patients at SR have a

higher rate of high severity ADEs compared to patients at EG and that patients in the ICU have a lower rate of D+ ADEs compared to those in general care. Similarly, patients hospitalized during T2, T3, T4 or T5 had a lower rate of high severity ADEs compared to baseline. The effect of the predictors can be verified in the bivariate plots from Figure 3.11.

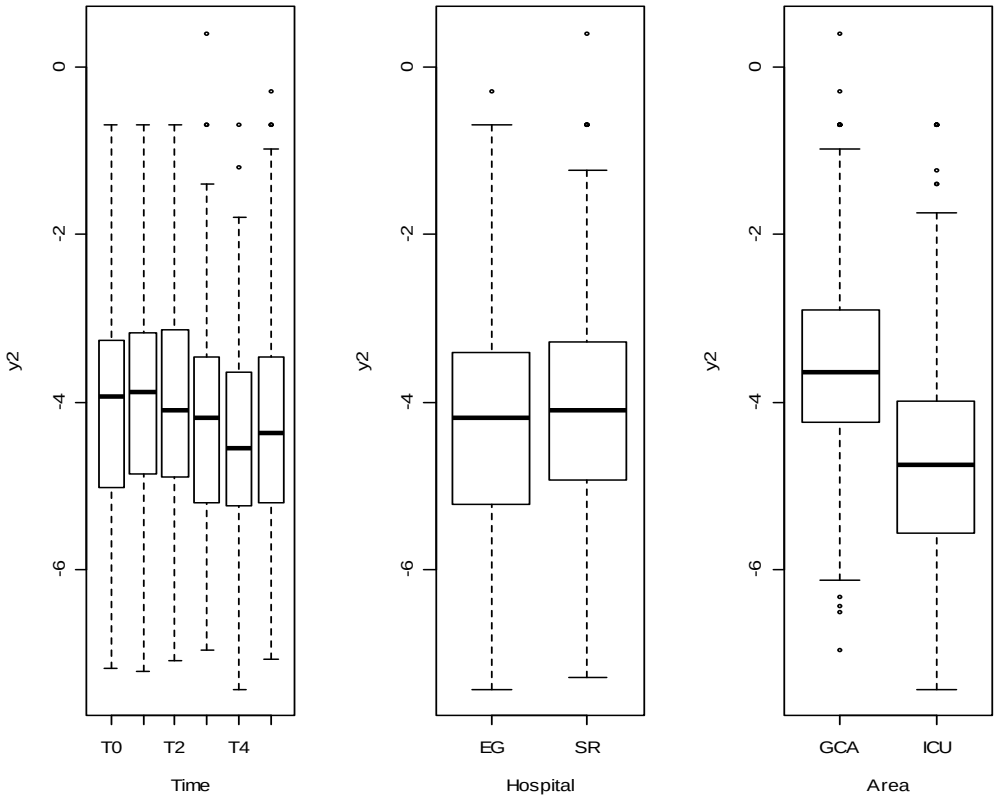

Figure 3.11 Bivariate plots of rate of high severity ADEs for each predictor
y2=log(ADEs+0.01)-log(Doses

We did not include interaction terms in the regression models for ADEs. However, based on the plots of ADEs over time in Figure 3.5 and Figure 3.6, it might be a good idea to consider and evaluate interactions between time period and area and between hospital and area.

A summary of the final models for each outcome measure is presented in Table 3.18. Despite the high number of zeros observed for the overall medication errors and ADEs, a zero-inflated model was not necessary, and a conditional Poisson model (conditional on the observed covariates) was enough to describe the observed occurrence of events. Warton et al [111] obtained similar results using abundance data. On the other hand, high-category events were better described by the zero-inflated Poisson. In this case, a zero-inflated model worked because the outcome measure only included those events that reached the patient and had the potential to cause harm. Thus, we can argue that, for instance, patients with mild conditions or patients of certain age are not at risk of ever experiencing those types of medication errors or ADEs.

Table 3.18 Summary of models for errors and ADEs

Outcome measure	**Best-fitting model**
Medication errors – all categories	Poisson
Medication errors - categories D and above	Zero-inflated Poisson
Adverse drug events	Poisson
Adverse drug events – categories D and above	Zero-inflated Poisson

Study Limitations

There were a few limitations to this study. First, because of the nature of the EMR implementation, it was not possible to isolate and evaluate the individual impact of the EMR functionalities (except for the EPICRx, since it was the first one). Therefore, the analysis performed only assessed the accumulated effect after each implementation (e.g., at T2, we evaluated the impact of EpicRx and eMAR together). In addition, other safety

and quality initiatives as well as building expansions that may have taken place during the EMR implementation were not included in this study. Thus, their potential effects on our outcome measures were not considered

Second, in several cases, it was difficult to establish whether a trigger indicated the occurrence of an ADE or not. Whenever a reviewer wasn't sure about the causality of an adverse event, it was written off as "Routinary" or "No event". This may be one reason for the low number of ADEs in the data.

Third, the use of number of doses as the exposure variable in our models was based on the assumption that every dose had the same probability of an error or ADE happening. However, in some cases, a number of doses might be administered at once (e.g., multi-dose vial or several pills taken at once), which alters the risk of an individual dose. Thus, a better way of defining and reporting the rate of events might have been using the number of orders, number of patients, patient-days, as seen in other similar studies [66, 69].

3.4 Conclusion

We have evaluated the impact on patient safety of the various functionalities of an EMR system in an inpatient pediatric setting. In this study, improvements in patient safety were represented by reductions in the occurrence of medication errors and adverse drug events. Our results indicated that the physician-related functionalities of the EMR (physician documentation and CPOE) had the largest effect in decreasing the rate of medication errors, which underscores the pivotal role of getting physicians to accept and efficiently use technologies such as CPOE. In the case of ADEs, the results suggested

that the main contribution of HITs in improving patient safety was in the reduction of the initial differences in rates between hospitals and between care areas. These variations may have been caused by inconsistencies in the various phases of the medication process among settings due to their different organizational cultures. By reducing sources of variation, it becomes easier to control and improve outcome measures [112].

In the literature review, we found that most healthcare studies examining the rates of medication errors and adverse drug events were descriptive in nature. Those attempting to model the occurrence or rate of errors or ADEs using statistical analysis mainly used logistic regression, which, depending on the underlying distribution of the data, might not be appropriate (e.g., if extra zeros are present). If the inappropriate distribution is assumed, there is the potential for drawing misleading inferences. Our study has contributed to the field by comparing six different nonlinear regression models for analyzing medication errors and adverse drug events in a pediatric inpatient facility.

When modeling all categories of errors and ADEs, the conditional Poisson (conditional on covariates) was the most adequate. The variances of both outcome variables were less than double their means, so they weren't overly dispersed (i.e., no need of NB). In addition, the high number of zeros did not appear to be due to extra zeros compared to a Poisson distribution, i.e., there was no evidence that the zeros were a mixture of two different types. For the case of medication errors and ADEs that reached the patient and either harmed him or required an intervention to avoid harm, the ZIP model was the best fit. The variances of these observed outcomes were also less than double their means, indicating that they weren't overly dispersed. So again, negative binomial models were not necessary. Also, since we were trying to model a specific

group of medication errors and ADEs, it was plausible to have a group of patients that never experienced those types of high severity events (two types of zeros). In summary, for count data with a high number of zeros, if the variance of the outcome variable is less than double its mean, then a Poisson model with covariates would be the best fit. In addition, if there is apriori knowledge of the potential presence of a subgroup where events never occur, then the ZIP model should also be tested.

3.5 Areas for future research

Areas for future work include:

- Evaluation of additional covariates such as length of stay and age and two-way interaction terms, in particular for ADEs, which presented a lot of variability between hospitals and care areas over time. Along the same lines, evaluation of a different set of covariates for the logistic part of the zero-inflated models in order to identify the at-no-risk populations.
- Exploration of other variables such as number of orders, patient-days to be used as the exposure variable in the count models, and incorporation of number of doses as a standard covariate.
- Investigation of the application of a Bayesian approach to calculate the coefficients of the parameter estimates, in particular for the models in which the parameters could not be estimated by maximum likelihood procedures.
- Evaluation of models that take into account covariates that might depend on time, i.e., inhomogeneous processes where the rate parameter is a function of time.

CHAPTER 4 : IDENTIFICATION OF PREDICTORS OF POSTOPERATIVE NAUSEA AND VOMITING

4.1 Literature Review

4.1.1 Post Operative Nausea and Vomiting

Post operative nausea and vomiting (PONV) is a condition that affects patients after undergoing surgery. It comprises three main symptoms (i.e., nausea, vomiting, and retching) that may occur separately or in combination. Nausea is the sensation of an urge to vomit; vomiting or emesis is the forcible expulsion of gastric contents through the mouth; and retching is an unproductive effort to vomit [113]. Although it rarely causes severe complications, patients are reportedly more afraid of PONV than of post operative pain [114]. Other consequences of PONV include increases in medical costs due to delays in patient discharge (Habib et al reported a mean delay of 25 min in PACU discharge [113]), additional nursing care in the inpatient setting, and unanticipated hospital admissions in the ambulatory setting [115-118].

The incidence of PONV is higher for general anesthesia than for regional anesthesia [114], however, some procedures such as cesarean section or major orthopedic surgeries have presented high rates of PONV after using regional anesthetics [118]. Despite our better understanding of PONV and the development of new anti-emetic drugs, the overall incidence of PONV remains between 20 and 30%. For high-risk groups, the incidence can be as high as 70% [116]. PONV can also occur following

discharge from post-anesthesia care units (PACU), even if the patient did not experience PONV in PACU [113].

4.1.2 Risks factors for PONV

PONV has a multifactorial origin. Independent risk factors for PONV can be classified as follow [114]:

- Patient-related factors: Well established factors are: gender, smoking status (not just smokers, but tobacco users in general [119]), history of PONV or motion sickness, and age. The incidence of PONV increases in female patients, in patients with a history of PONV or motion sickness and decreases with age (for adults) [117]. Factors previously considered, but that have been disproved are body mass index and early stage of menstruation cycle [120, 121].
- Surgical-related factors: type and duration of surgery (higher risk for some procedures) are indicated as risk factors for PONV, although there is controversy regarding their independence [121, 122]
- Anesthesia-related factors: administration of volatile anesthetics, opioids, N_20 or neostigmine can increase the incidence of PONV. Likewise, longer duration of anesthesia can increase PONV [117, 123]

Another possible predictor is the physical status classification by the American Society of Anesthesiologists (ASA) [121]. This is a system for assessing the fitness of patients before selecting the anesthetic or performing surgery, but is not intended to predict operative risk. It is used by anesthesiologists to classify patients according to their medical history. The categories of the ASA system are described in Table 4.1. The genetic background of a patient as a risk factor is also starting to be investigated [124].

Table 4.1 ASA physical status classification system

Category	Description
ASA 1	A normal healthy patient
ASA 2	A patient with mild systemic disease
ASA 3	A patient with severe systemic disease
ASA 4	A patient with severe systemic disease that is a constant threat to life
ASA 5	A moribund patient who is not expected to survive without the operation
ASA 6	A declared brain-dead patient whose organs are being removed for donor purposes

Depending on the surgeries and the populations covered in the reviewed studies, the risk factors can vary significantly from one study to another [125]. However, the following risk factors have been reported in most studies [126]:

- gender
- history of PONV or motion sickness
- duration of anesthesia
- non-smoking status
- use of opioids

In most studies of PONV, the outcome analyzed is either vomiting or nausea and vomiting together. However, nausea and vomiting have different physiopathologies, which mean that they may have different risk factors. For instance, Stadler et al found that gender, nonsmoking status, and type of anesthesia were predictive of both nausea and vomiting, but type of surgery was associated with nausea but not vomiting [127].

4.1.3 Management of PONV

New substances have been developed and successfully introduced for the prevention and treatment of PONV. These include NK1 receptor antagonists, new 5-HT_3

receptor antagonists, dimenhydramin, droperidol, dexamethasone, among others [114, 117, 128]. In order to increase the antiemetic efficacy of single interventions, several studies have compared the efficacy of combinations of antiemetic regimens and multimodal use of anesthetic techniques [117]. The results show that a combination of interventions is more effective than using a single agent (combining interventions has an additive effect in risk reduction) [129]. Furthermore, the use of multimodal approaches (pharmacologic and nonpharmacologic) to reduce PONV offers additional advantages over drug combinations [128, 130].

Guidelines for the management of PONV have been developed in the last decade. Numerous studies and panels of experts agree that a risk-based approach is the most effective strategy for the prevention and treatment of PONV, and can reduce the overall institutional rate of PONV [131]. Antiemetics are not universally administered because it is not a cost-effective practice and because of their additional adverse effects [132], which include increased risk of post operative infection [133]. According to the risk-based approach, after an initial estimation of the patient's risk for PONV, only those at moderate and high risk should be administered antiemetic prophylaxis. Interventions to reduce baseline risk factors should be used first (e.g., avoid inhalational anesthesia which has been shown to cause early postoperative vomiting [123]). Depending on the level of risk, either a monotherapy or a combination therapy is indicated. Basically, the higher the risk of PONV, the greater the number of antiemetic interventions that should be used is [134, 135]. A multimodal approach with combinations of antiemetics and nonpharmacologic approaches should be considered for patients at high risk of PONV. For rescue treatments, the antiemetic administered should be from a different class than

the ones used for prophylaxis [136-139]. In Figure 4.1, we can observe the process for the risk-based approach.

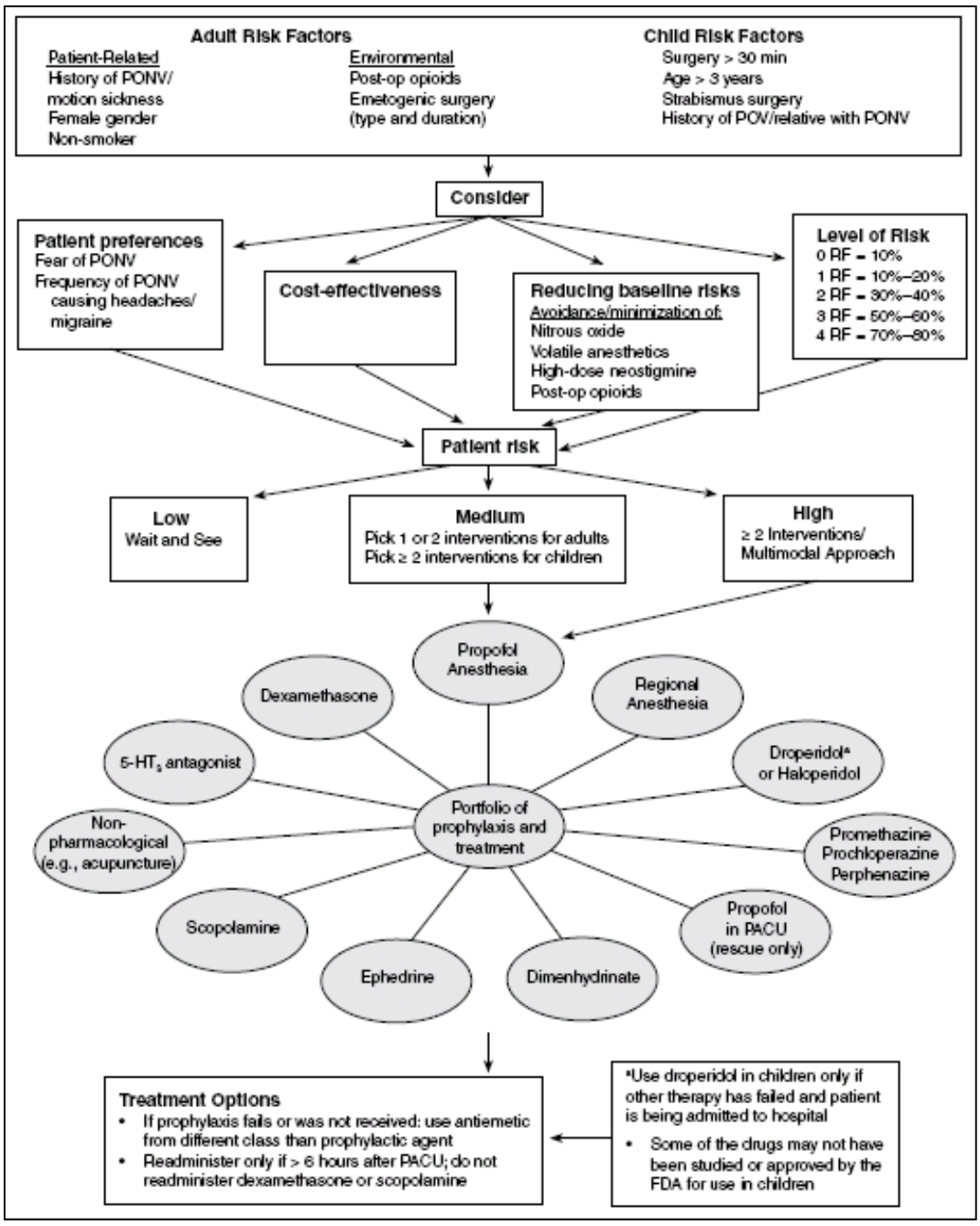

Figure 4.1 Algorithm for management of PONV [136]

4.1.4 Predictive models for PONV

A number of predictive models and simplified scoring systems to quantify the likelihood of PONV have been developed. Most of them have used logistic regression analysis to identify the risk factors for PONV; although there have also been attempts at different methodologies such as decision trees [140] and neural networks [141], the latter with promising results. Among the most common scores are the Apfel, Koivuranta, Palazzo and Evans, and Sinclair scores [121].

The formula of the scoring systems for the probability of PONV is:

$$P = \frac{1}{1+e^{-S_{coeff}}} \tag{4.1}$$

$$S_{coeff} = b_0 + b_1x_1 + b_2x_2 \ldots + b_kx_k \tag{4.2}$$

where P is the probability that a patient will have PONV and $-S_{coeff}$ is the logistic regression equation with the coefficients b_i for each predictor variable x_i

In the Apfel model, the following predictors were identified: gender, history of motion sickness or PONV, smoking and use of post operative opioids. The type of surgery was not a relevant factor in predicting PONV.

$$y = -2.28 + 1.27\, female\ gender + 0.65\, history\ of\ PONV\ or\ motion\ sickness + 0.72\, nonsmoking + 0.78\, postoperative\ opioid\ use \tag{4.3}$$

A simplified version of this score lets clinicians quantify the risk for PONV by counting the number of risk factors present: 10% rate for 0 risk factors, 23% for 1 factor, 39% when 2 are present, 61% when 3 factors are present and 79% when all 4 factors are present [142]. These results have been confirmed by other studies, in which the risk scoring system was highly predictive of emetic symptoms within 24 h after surgery (even

after multimodal antiemetic prophylaxis); however no correlation was found between the Apfel risk score and the incidence of late PONV (> 24 h after surgery) [143].

The Koivuranta scoring system [144] included the following predictors: gender, history of PONV, history of motion sickness, non-smoker, and duration of surgery over 60 min:

$$y = -2.21 + 0.93\, female\ gender + 0.82\, history\ of\ PONV$$
$$+0.75\, duration\ of\ surgery\ over 60\,\text{min} + 0.61\, nonsmoking$$
$$+ 0.59\, history\ of\ motion\ sickness \quad (4.4)$$

A simplified version of the Koivuranta score gives the following prediction probabilities: 0.17, 0.18, 0.42, 0.54, 0.74, and 0.87 for the presence of 0 to 5 risk factors.

A study evaluating the Apfel and the Koivuranta scores found that they do not provide accurate prediction of the risk of PONV in populations different to the one used to develop the scores [145]. Similarly, Thomas et al reported lack of agreement between scoring systems [146]. Comparisons of scoring systems have demonstrated that simplified scores are easier to use [147], and provide equal or better discriminating power than more complex ones [148-151]. Recommended scoring systems for adults are the Apfel and the Koivuranta ones, while the Eberhart simplified system is recommended for children [121].

The Palazzo & Evans model [152] is effective at estimating the risk of PONV in groups of patients, particularly those at high risk [153].

$$y = -5.03 + 2.34\, post\ operative\ opioid\ use + 3.97\, history\ of\ PONV$$
$$+2.4\, female\ gender + 0.78\, history\ of\ motion\ sickness$$
$$-3.2\, female\ gender\ and\ previous\ PONV \quad (4.5)$$

The Sinclair score [154] identified the following independent predictors: sex, age, smoking status, previous PONV, type of anesthesia used, duration of anesthesia, and type of surgery. Women and non-smokers had three times more probability of developing PONV than men and smokers, respectively. A 10-year increase in age decreased the likelihood of PONV by 13%, but only for those older than 50 yr. A 30-min increase in the duration of anesthesia increased the likelihood of PONV by 59%. Likewise, general anesthesia increased the likelihood of PONV 11 times more than other type of anesthetic technique.

In her master's thesis, Velickovic [155] developed a score with three separate regressions models for each time interval (0-2, 2-24, 0-24). The factors included in the models were: type of surgery (parateoidectomy), history of PONV, BMI, smoking and gender.

$$[0-2h]: \quad y = -0.964 - 2.306\,para + 0.938\,PA \tag{4.6a}$$

$$[2-24h]: \quad y = -4.482 + 1{:}514\,PA + 0.062\,BMI + 1.492\,gender \tag{4.6b}$$

$$[0-24h]: \quad y = -1.627 - 1.366\,para + 1.417\,PA + 0.525\,smoking + 0.864\,gender \tag{4.6c}$$

A summary of the currently used PONV models is presented in Table 4.2. We observed that the number as well as the type of predictors are different for each model and tend to be specific to the dataset used to develop the model (e.g., type of surgery in the LELA model). Thus, the models are not robust for all populations.

Table 4.2 Predictors for five common PONV models

Models	Apfel	Koivuranta	Palazzo & Evans	Sinclair	LELA
# predictors	4	5	5	7	4
Predictors	female gender	female gender	female gender	female gender	female gender
	history of PONV or motion sickness	history of PONV	history of PONV	history of PONV	history of PONV
	nonsmoking	history of motion sickness	history of motion sickness	smoking status	smoking status
	PO opioid use	nonsmoking	PO opioid use	age	type of surgery (parathyroidectomy)
		duration of surgery over 60 min	female gender x history of PONV	type of anesthesia used	
				duration of anesthesia	
				type of surgery	

4.1.5 Summary of key research needs and study aims

From the literature review, we extracted the following points:

- Although several scoring systems have been developed in the last decade, the current scoring systems are not robust among populations. Either adjustment of the scoring system to the new population, or re-estimation of the regression coefficients is necessary to use a previously developed scoring system.
- Although several studies have proved that adding more risk factors will not increase the discriminating power of the current models, there is still more research to be done [156] and better antiemetic approaches are still needed [157].

In this study, we attempt to develop a more accurate but also more general model to predict the probability of PONV regardless of the type of surgery.

4.2 Methodology

The objective in this chapter is to develop a model to predict the risk of having PONV during the first 2 hours after surgery (early PONV) and from 2 to 24 hours after surgery.

4.2.1 Data Description

The data used for this study came from the database for anesthetic service of the Center for Endocrine Surgery in Belgrade, Serbia. It was a retrospective study in which 471 patients were investigated to identify risk factors for PONV. Patients in the study underwent some type of thyroidectomy procedure. Standard anesthetic technique was conducted in all patients. The patients were not given antiemetics either in the preoperative period or during the surgery. In the case of appearance of two or more instances of postoperative nausea and vomiting or after prolonged postoperative nausea, antiemetics were given. The monitoring system during the surgery included EKG, pulse oximetry, none invasive assessment of blood pressure, and capnography.

Outcome variables

The outcome measure was PONV, a binary variable defined as the existence of at least one episode of postoperative nausea, vomiting or nausea and vomiting together. As such, it was recorded two times, after 2 hours after surgery and from 2 to 24 hours after surgery.

Potential predictors

The following variables were investigated for predictive power:

- Variables related to the patient

- Age (yr)
- Gender
- History of PONV
- History of motion sickness
- ASA classification
- Smoking status
- Body mass index (kg/m2)

- Variables connected with anesthesia:
  - Duration of anesthesia (in min)
  - Number of attempts for intubation
  - Duration of intubation (in min)
  - Time from the end of anesthesia till extubation (in min)

4.2.2 Analysis Methods

An initial assessment to identify associations between predictors and the response variable was done using univariate analysis. The continuous variables were analyzed using the Student's t-test and the categorical variables using the χ^2 test (2x2 contingency tables). In an effort to simplify the interpretation of the final predictive models, some continuous variables were transformed into categorical ones. A p-value of <0.05 was considered significant.

Predictive models using logistic regression models were built for two outcome measures:

- Early PONV: Prediction of PONV during the first 2 hours after surgery
- Late PONV: Prediction of PONV from 2 to 24 hours after surgery

The categorical predictors were indicated as 1 if present, and 0 if absent (for gender, 1 was female, and 0 was male). The inclusion of factors was determined using backward stepwise regression; and a p-value less than 0.05 was used as the criterion for variables retention. Interaction effects were also examined.

Predictive accuracy of the model was assessed by the area under the curve (AUC) of the receiving operating characteristic (ROC) curve. Additionally, both PONV models were compared to previously developed risk scores: Apfel, Simplified Apfel, Sinclair and Lela, using the AUC. All statistical analyses were performed using SPSS Statistics v19.0 and MATLAB (R2010b).

4.3 Results and Discussion

Of the 471 patients, five were excluded because they belonged to category 6 in the ASA classification (brain-dead patients). Only patients in category 5 or less were included in this study. A total of 466 patients on whom the antiemetic prophylaxis was not applied were assessed.

Patients' age varied from 12 to 81 yr (mean = 51.76, SD = 13.65). The female to male ratio was 4:1, and about one third of the patients were smokers. Other demographic and clinical characteristics of the patients, as well as the distribution of the categorical variables are described in Table 4.3 and Table 4.4.

Table 4.3 Descriptive statistics for continuous variables

Variable	N		Mean	Std. Deviation	Median	Min	Max
	Valid	Missing					
Age (yr)	466	0	51.76	13.65	54.00	12.00	81.00
Duration of anesthesia (min)	466	0	78.20	25.84	75.00	20.17	240.00
Weight (kg)	466	0	73.90	15.39	72.00	38.00	145.00
BMI (kg/m2)	466	0	26.11	4.64	25.46	15.00	42.39
Time until extubation (min)	464	2	19.22	133.91	10.00	0.00	2880.00

Table 4.4 Distribution of categorical variables (n=466)

Variable		Frequency	Percent
Gender	Male	91	19.5
	Female	375	80.5
Smoking Status	Smoker	158	33.9
	Non smoker	308	66.1
History of Motion Sickness	No	440	94.4
	Yes	26	5.6
History of PONV	No	409	87.8
	Yes	57	12.2
Surgery Type	Lobectomy	4	0.9
	Bilateral subtotal lobectomy	1	0.2
	Total thyroidectomy	300	64.4
	Hemithyroidectomy	92	19.7
	Parathyroidectomy	50	10.7
	Unilateral dissection of neck	5	1.1
	Other	14	3.0
ASA Classification	1	92	19.7
	2	1	0.2
	3	278	59.7
	4	1	0.2
	5	94	20.2
Intubation Trials	1	357	76.6
	2	74	15.9
	3	27	5.8
	4	7	1.5
	5	1	0.2
Duration of Intubation	1: Standard	429	92.1
	2: Extended	33	7.1
	3: More than 10 min	4	0.9

In order to simplify the use and interpretation of the predictive models, some continuous variables were transformed to binary variables (Table 4.5). Based on the risk factors identified in previous PONV studies and scores, the following variables were added: age over 50, duration of anesthesia over 75 min, BMI over 30 (or obese). Likewise, ASA physical status classification was reduced to only 2 categories: one including the original categories 1, 2 and 3 (better health status) and another one for categories 4 and 5.

Table 4.5 Distribution of new categorical variables (n=466)

Variables		Frequency	Percent
Age over 50	No	189	40.6
	Yes	277	59.4
BMIGR2	Not obese	384	82.4
	Obese	82	17.6
Duration of anesthesia over 75 min	No	252	54.1
	Yes	214	45.9
ASAclass2	0: Cat 1-3	371	79.6
	1: Cat 4-5	95	20.4
Duration Intubation2	Standard	429	92.1
	Extended	37	7.9

More patients experienced early PONV than late PONV. More than a quarter of the patients (27.3%) had at least one episode of PONV during the first two hours after surgery, while only 21.5% experienced PONV during the next 22 hours (Table 4.6).

Table 4.6 Observed incidence of PONV (n=466)

Outcome Variable		Frequency	Percent
PONV within 2 hr after surgery	No	339	72.7
	Yes	127	27.3
PONV from 2 to 24 hr after surgery	No	366	78.5
	Yes	100	21.5

4.3.1 Exploratory analysis of predictors

Continuous predictors

Continuous variables were evaluated using the t test. No significant differences between patients with and without PONV episodes (for both early and late PONV) were identified at $\alpha=0.05$. However, for late PONV, duration of anesthesia and BMI had p-values of 0.06 and 0.078, respectively, and thus were included as initial predictors in the logistic regression models (Table 4.7).

Table 4.7 Results from t-test for continuous variables

Group		Age	Duration of anesthesia	Weight	BMI	Time until extubation
PONV 0 to 2	t	0.977	-0.088	1.461	-0.032	0.593
	df	464	312.516	464	464	462
	p-value	0.329	0.93	0.145	0.975	0.554
PONV 2 to 24	t	-0.878	-1.884	0.401	-1.767	0.468
	df	464	464	464	464	462
	p-value	0.38	0.06	0.689	0.078	0.64

Categorical Predictors

The Pearson Chi-square test for independence between binary categorical predictors and early PONV was statistically significant for gender ($\chi^2=5.334$, $p=0.025$), history of PONV ($\chi^2=13.254$, $p<0.001$), and ASA classification ($\chi^2=6.524$, $p=0.014$) (Table 4.8). Similarly, statistically significant associations were identified between late PONV and the following categorical variables: gender ($\chi^2=12.716$, $p<0.001$), history of PONV ($\chi^2=25.866$, $p<0.001$), duration of anesthesia over 75 min ($\chi^2=10.16$, $p=0.002$), and early PONV ($\chi^2=27.642$, $p<0.001$) (Table 4.9).

All categorical variables with p value less than 0.5 were included as initial predictors in the logistic regression models for both early and late PONV.

Table 4.8 2x2 Contingency tables for early PONV

Variable		PONV0to2		Total	Pearson Chi-Square	p-value
		no	yes			
gender	male	75	16	91	5.334	**.025**
	female	264	111	375		
	Total	339	127	466		
ageover55	no	178	69	247	0.123	.755
	yes	161	58	219		
	Total	339	127	466		
BMIGR2	Not obese	277	107	384	0.411	.586
	Obese	62	20	82		
	Total	339	127	466		
Smoking	yes	121	37	158	1.774	.189
	no	218	90	308		
	Total	339	127	466		
HistMS	no	320	120	440	0.002	1.000
	yes	19	7	26		
	Total	339	127	466		
HistPONV	no	309	100	409	13.254	**.000**
	yes	30	27	57		
	Total	339	127	466		
Duranesover75	no	189	63	252	1.405	.252
	yes	150	64	214		
	Total	339	127	466		
ASAclass2	0: Class 4-5	79	16	95	6.524	**.014**
	1: Class 1-3	260	111	371		
	Total	339	127	466		
DurIntub2	0: Standard	315	114	429	1.259	.335
	1: Extended	24	13	37		
	Total	339	127	466		

Table 4.9 Contingency tables for late PONV

Variable		PONV2to24		Total	Pearson Chi-Square	p-value
		no	yes			
gender	male	84	7	91	12.716	**0.000**
	female	282	93	375		
	Total	366	100	466		
ageover55	no	196	51	247	0.205	0.653
	yes	170	49	219		
	Total	366	100	466		
BMIGR2	Not obese	307	77	384	2.564	0.137
	Obese	59	23	82		
	Total	366	100	466		
Smoking	yes	131	27	158	2.709	0.121
	no	235	73	308		
	Total	366	100	466		
HistMS	no	345	95	440	0.081	0.815
	yes	21	5	26		
	Total	366	100	466		
HistPONV	no	336	73	409	25.866	**0.000**
	yes	30	27	57		
	Total	366	100	466		
duranesover75	no	212	40	252	10.16	**0.002**
	yes	154	60	214		
	Total	366	100	466		
ASAclass2	0: Class 4-5	80	15	95	2.276	0.161
	1: Class 1-3	286	85	371		
	Total	366	100	466		
DurIntub2	0: Standard	336	93	429	0.154	0.836
	1: Extended	30	7	37		
	Total	366	100	466		
Headache0to2	no	335	90	425	0.229	0.690
	yes	31	10	41		
	Total	366	100	466		
PONV0to2	no	287	52	339	27.642	**0.000**
	yes	79	48	127		
	Total	366	100	466		

4.3.2 Logistic Regression models

Model for early PONV (0 to 2 hr after surgery)

The following predictors were evaluated as independent variables in the model for early PONV: gender, smoking, HistPONV, duranesover75, ASAclass2, DurIntub2 and BMI. The following variables met the criteria for inclusion when alpha was set to 0.05: Gender, ASA physical status, and history of PONV. From Table 4.10, we observed that the odds of developing an episode of PONV during the first 2 hours after surgery was almost double for women that for men (OR=1.768). Likewise, the odds for patients with history of PONV were 2.477 times higher than for those with no previous history of PONV. Lastly, a patient with a better physical status (ASA 1-3), had about twice the odds (OR=1.892) of developing early PONV that one classified in either category 4 or 5.

Table 4.10 Logistic regression model for early PONV - 3 predictors

Variable	B	S.E.	Wald	df	Sig.	Exp(B)	95% CI for Exp(B)	
							Lower	Upper
Gender female	0.570	0.303	3.534	1	0.060	1.768	0.976	3.203
Better physical status	0.638	0.302	4.471	1	0.034	1.892	1.048	3.418
HistPONV	0.907	0.292	9.620	1	0.002	2.477	1.396	4.395
Constant	-2.097	0.377	30.850	1	0.000	0.123		

When the inclusion criteria α was increased to 0.1, patients' smoking status (variable Nonsmoking) was also included in the model (see Table 4.11). The odds ratio for nonsmokers was 1.5 times higher than for smokers. The odds and coefficients for the previously selected predictors changed minimally.

Table 4.11 Logistic regression model for early PONV - 4 predictors

Variable	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
							Lower	Upper
Gender female	0.565	0.304	3.452	1	0.063	1.760	0.969	3.196
NonSmoking	0.407	0.234	3.030	1	0.082	1.502	0.950	2.376
HistPONV	0.903	0.294	9.418	1	0.002	2.466	1.386	4.389
Better physical status	0.683	0.304	5.047	1	0.025	1.979	1.091	3.590
Constant	-2.408	0.424	32.322	1	0.000	0.090		

Several models with different 2-way interaction terms were also evaluated. The interaction terms Gender*Duranover75, Gender*Ageover55 and Ageover55*HistPONV were significant at $\alpha = 0.1$ (one interaction effect per model). Only the model including the interaction Gender*Ageover55 (Table 4.12) also improved on the value of the AUC.

Table 4.12 Logistic regression model for early PONV – 5 predictors

Variable	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
							Lower	Upper
Gender female	0.072	0.363	0.039	1	0.843	1.075	0.528	2.189
HistPONV	0.937	0.295	10.094	1	0.001	2.551	1.432	4.546
Better physical status	0.694	0.315	4.861	1	0.027	2.002	1.080	3.710
Ageover55	-1.291	0.690	3.503	1	0.061	0.275	0.071	1.063
Female*over55	1.520	0.728	4.368	1	0.037	4.574	1.099	19.038
Constant	-1.772	0.423	17.514	1	0.000	0.170		

The values of the AUC for the models with 3, 4 and 5 predictors were 0.617, 0.634 and 0.642, respectively. The ROC curves for each model are displayed in Figure 4.2. Based solely on the AUC, the model with 5 predictors would be the best option. However, many times we are also interested in the model performance above a minimum sensitivity or specificity value. In this particular study, and to simplify the calculations, we assigned minimum sensitivity and specificity values of 60% and 30%, respectively.

Assuming that sensitivity and specificity are equally important, the model with the optimal sensitivity-specificity combination would be the one with the largest distance to the reference line.

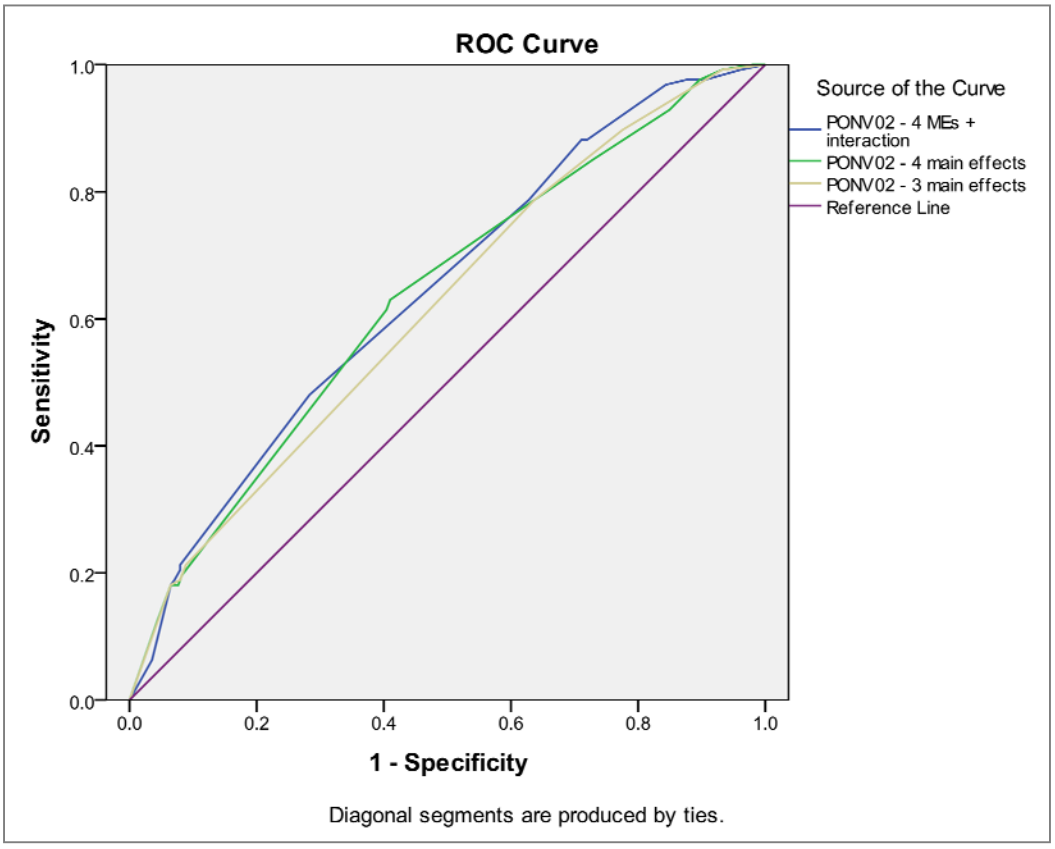

Figure 4.2 ROC curves for early PONV models with 3, 4 and 5 predictors

For each curve, its AUC and the coordinate with the largest distance to the reference line, within the specified sensitivity-specificity range, are presented in Table 4.13. The model with 4 predictors had the largest distance, and was therefore selected as the best model to predict early PONV.

Table 4.13 Best sens-spec combinations for sensitivity >= 60%

Model	AUC	Cutoff value	Sensitivity	Specificity	Distance to reference line
3 predictors	0.617	0.24	0.787	0.363	0.106
4 predictors	**0.634**	**0.26**	**0.630**	**0.590**	**0.155**
5 predictors	0.642	0.26	0.787	0.372	0.112

The final model for early PONV can be expressed as:

$$y = log\left(\frac{p}{1-p}\right) = -2.408 + 0.565 * gender + 0.407 * Nonsmoking$$

$$+0.683 * ASAclass2inv + 0.903 * HistPONV$$

Where p = probability of having an episode of PONV during the first 2 hr after surgery, is:

$$p = \frac{0.09 * 1.76^{gender} * 1.502^{Nonsmoking} * 1.979^{ASAclass2inv} * 2.466^{HistPONV}}{1 + 0.09 * 1.76^{gender} * 1.502^{Nonsmoking} * 1.979^{ASAclass2inv} * 2.466^{HistPONV}}$$

Models for late PONV (2 to 24 hr after surgery)

The following predictors were included in the final model: gender female, previous history of PONV, whether they had PONV during the first 2 hr after surgery, duration of anesthesia over 75 min and body mass index (in Table 4.14). The presence of each predictor increases the likelihood of having late PONV. For BMI, the odds increase in 5% per each unit increase. The AUC for the ROC curve of this model is 0.75.

Another model using the categorical representation of BMI (BMIGR2), which differentiates obese vs. not obese patients, was also evaluated. Additionally, and in an effort to make the model more applicable to other populations, we replace the categorical variable 'Duranesover75' for the continuous predictor 'Duration of anesthesia'. The results of this model are presented in Table 4.15. The AUC for this model was 0.738.

An even simplified version without the predictor 'duration of anesthesia' was also examined (Table 4.16). The AUC for this model was 0.728.

Table 4.14 Model for late PONV (4 cat pred + BMI)

Variable	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for Exp(B)	
							Lower	Upper
Gender female	1.379	.451	9.347	1	.002	3.969	1.640	9.607
HistPONV	1.204	.322	13.978	1	.000	3.333	1.773	6.265
PONV0to2	1.012	.253	15.989	1	.000	2.750	1.675	4.516
DurAnesover75	.615	.249	6.114	1	.013	1.849	1.136	3.009
BMI	.051	.027	3.692	1	.055	1.053	.999	1.109
Constant	-4.702	.863	29.652	1	.000	.009		

Table 4.15 Model for late PONV (4 cat pred + DurAnesthesia)

Variable	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
							Lower	Upper
Gender female	1.410	.452	9.730	1	.002	4.097	1.689	9.939
HistPONV	1.265	.320	15.628	1	.000	3.543	1.892	6.633
PONV0to2	1.037	.253	16.859	1	.000	2.821	1.719	4.627
DurAnesthesia	.009	.005	3.841	1	.050	1.009	1.000	1.019
BMIGR2	.594	.307	3.738	1	.053	1.811	.992	3.305
Constant	-3.945	.615	41.096	1	.000	.019		

Table 4.16 Model for late PONV with 4 categorical predictors

Variable	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for Exp(B)	
							Lower	Upper
Gender female	1.355	.450	9.081	1	.003	3.876	1.606	9.357
HistPONV	1.287	.318	16.433	1	.000	3.623	1.944	6.751
PONV0to2	1.037	.251	17.087	1	.000	2.819	1.725	4.609
BMIGR2	.661	.304	4.725	1	.030	1.937	1.067	3.515
Constant	-3.165	.450	49.572	1	.000	.042		

Figure 4.3 presents the ROC curves for all three models for late PONV.

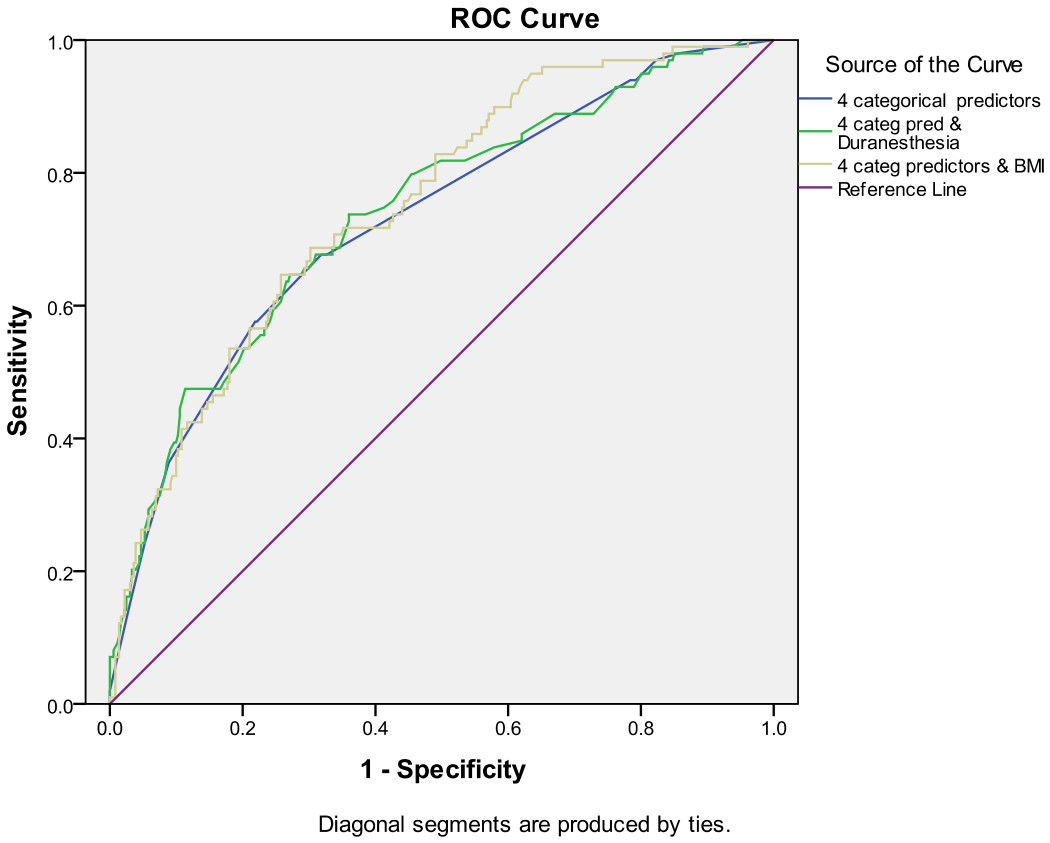

Figure 4.3 ROC curves for models of late PONV

In addition to the AUC for each models, we also identified the model with the best sensitivity-specificity combination for a particular range. In this case, since the values of AUC are higher than for early PONV, we assigned a minimum value of 65% for both sensitivity and specificity. We calculated the largest distance to the reference line for each model, and obtained their optimal sensitivity-specificity combinations. The results are presented in Table 4.17. We observed that there is not much difference among the distances of the models; thus, in the interest of parsimony, we selected the model with 4 predictors as the optimal predictive model for late PONV.

Table 4.17 Best sens-spec combinations for sensitivity >= 65%

Model	AUC	Cutoff value	Sensitivity	Specificity	Distance to reference line
4 cat predictors	0.728	0.23	0.677	0.681	0.253
4 cat predictors + Duranes	0.738	0.20	0.677	0.690	0.259
4 cat predictors + BMI	0.750	0.21	0.687	0.698	0.272

The final model for late PONV can be expressed as:

$$y = log\left(\frac{p}{1-p}\right) = -3.165 + 1.355 * gender + 0.661 * BMIGR2$$

$$+1.287 * HistPONV + 1.037 * PONV0to2$$

Where p = probability of having at least one episode of PONV between 2 and 24 hr after surgery:

$$p = \frac{0.042 * 3.876^{gender} * 1.937^{BMIGR2} * 3.623^{HistPONV} * 2.819^{PONV0to2}}{1 + 0.042 * 3.876^{gender} * 1.937^{BMIGR2} * 3.623^{HistPONV} * 2.819^{PONV0to2}}$$

Comparison with other Scores

Both the early and late PONV models were compared to the LELA scores, which were developed using the same patient database.

The AUC values and ROC curves for the early PONV models are presented in Table 4.18 and Figure 4.4. The model with 4 predictors has the highest AUC. The Lela score has the advantage of having only 2 predictors; however, one of them is surgery-related and specific to the population used to develop the score. Thus, it makes it less applicable to use with other populations. Our model, although with two more predictors, uses general categorical patient-related variables which can be easily calculated to

determine the patient risk for PONV. Additionally, it's easy to see that the model with 4 predictors has the best sensitivity- specificity combination for sensitivity > 60%.

Table 4.18 Comparison of early PONV models with LELA score

Model	AUC
PONV0to2 - 4 predictors	0.634
PONV0to2 - 3 predictors	0.617
Lela Score for PONV0to2	0.615

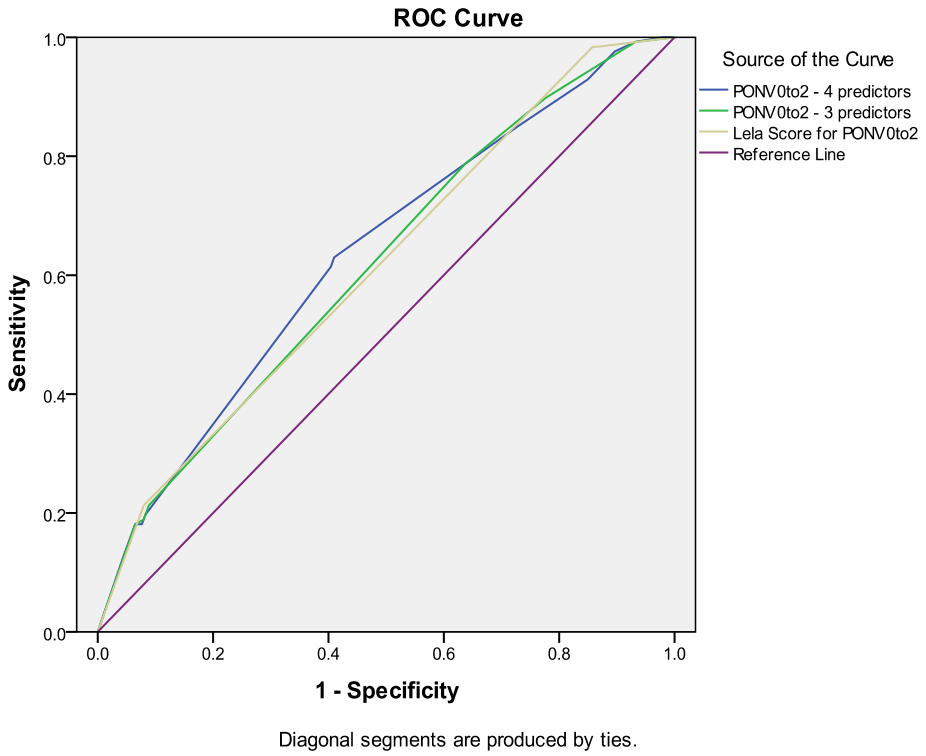

Figure 4.4 ROC curves comparing models for early PONV vs. Lela score

Additionally, our model for early PONV was also compared to other well known PONV scores. The model with 4 predictors was better than the Sinclair, Apfel and the simplified Apfel. The AUC values and ROC curves for all the models are shown in Table 4.19 and in Figure 4.5.

Table 4.19 AUC for models of early PONV

Model	AUC
PONV0to2 - 4 predictors	**0.634**
Sinclair calculated probability	0.560
Apfel calculated probability	0.595
Apfel Score	0.597

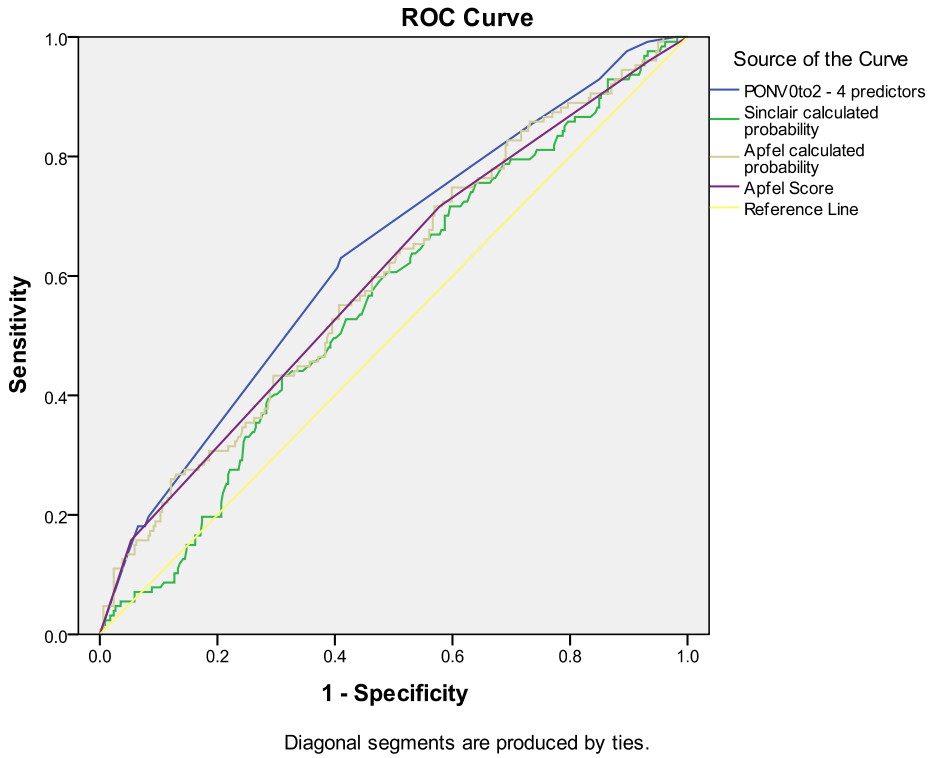

Figure 4.5 ROC curves comparing early PONV model to Apfel & Sinclair scores

In a similar fashion, the selected regression model for late PONV was also compared with the LELA, Apfel, Sinclair and simplified Apfel scores. The results in Table 4.20 show that our model is the best one with the highest AUC value, followed by the LELA score. The Sinclair and Apfel scores have AUC values lower than 65%. As in the case with early PONV, the LELA score for late PONV has fewer predictors (3), but

includes a categorical variable for type of surgery and a continuous variable for BMI. Our model has 4 predictors, but they are all categorical, which makes it easier to develop a simplified score. Furthermore, the ROC curves in Figure 4.6 show that our model with 4 predictors also has the best sensitivity-specificity combination.

Table 4.20 AUC for models of late PONV

Model	AUC
Late PONV - 4 predictors	0.725
LELA score for PONV2to24	0.690
Sinclair calculated probability	0.602
Apfel calculated probability	0.637
Apfel Score	0.648

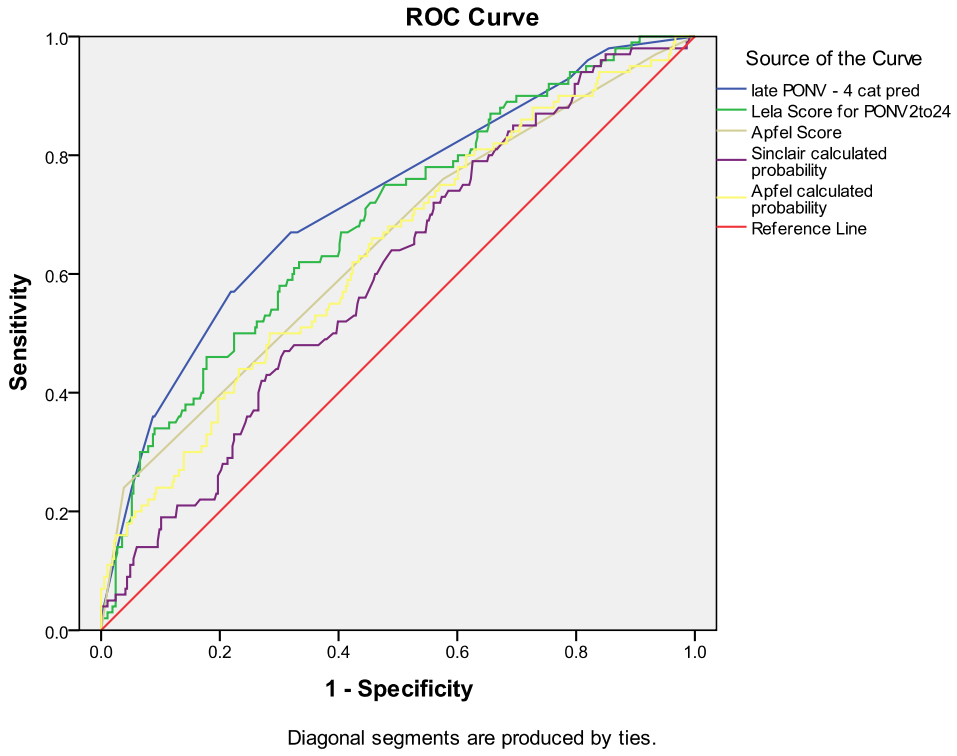

Figure 4.6 ROC curves comparing the model for late PONV to other scores

4.4 Conclusions

In both the early and late PONV models, the most important predictor of PONV was previous history of PONV. For early PONV, ASA physical status was also a significant predictor. We found only one study [121] that mentioned better ASA physical status as a possible risk factor for PONV. Our results make stronger the case for including the patient's ASA physical status as a risk factor for early PONV.

For late PONV, not only previous history of PONV was a significant predictor, but also early PONV. This is worth noting because if a patient did not have an episode of PONV during the first 2 hours, then is most likely that he/she would not experience PONV in the next 22 hours. Although BMI was discarded as a risk factor in several studies, our research found it to be a strong predictor of later PONV, with obese patients having less probability than not obese patients of developing PONV. Often cited risk factors such as age or length of surgery were not significant in our study.

Our model was stronger than previously developed scores. Another advantage of the developed models is that the significant predictors in both early and late PONV models are factors that can be easily assessed before surgery to obtain the patient probability of developing PONV.

Some limitations of this study were:

- The patients included in this study underwent some type of thyroidectomy. Thus, although we tried to make the model more general by not including the type of surgery, our model might not work as well for patients undergoing other type of surgeries.

- There was no record of the number of PONV episodes a patient experienced; therefore, we didn't have data on whether a patient was given antiemetics during the first 2 hours after surgery. This could have an effect on the significance of early PONV on later PONV that we did not consider.

Future work includes:

- Evaluate the robustness of the models and its independence from surgery type with an independent dataset
- Investigate the significance of early PONV in the late PONV with additional data on whether a patient was given antiemetics during the first 2 hours after surgery
- Develop a model for PONV during the entire 24 hours after surgery and also individual models for post operative nausea and post operative vomiting

CHAPTER 5 CONCLUSION

This thesis deals with the application of predictive modeling and data mining to analyze and interpret three types of data found in healthcare and clinical research. The first study involves the assessment of hydration status using breath analysis. Through analysis of the more than 300 volatile organic compounds contained in exhaled human breath, we proposed to identify markers of hydration. With a sample size of 36 subjects, this requires the use of methods to solve a "large p, small n" problem. The motivation behind this study is that current hydration assessment methods are either, accurate and invasive (e.g., urine osmolality), or not invasive and also not very accurate (e.g., change in body mass). On the other hand, breath tests are inexpensive, noninvasive and easy to perform, and would be a better alternative to current hydration measurement methods.

Breath test data consist of a large number of volatile organic compounds (VOCs), serving as potential predictors and biomarkers. A methodology to reduce the number of VOCs is developed by correlating each of them with a "hydration profile" based on the subjects' hydration probabilities. We further apply principal component analysis and support vector machines and build a classifier to discriminate between the euhydrated and the 24-hr dehydrated group. The classification rate is 73.89%, which is low when compared to two other well-known markers of hydration: U_{osm} and P_{osm}. The former provides almost a 100% classification rate, and the latter has 82%. Although our results do not provide an improvement on other hydration measures, this study was the first to explore the application of breath analysis to measure hydration status. Future research includes the identification of the VOCs associated to the dehydration process

and the evaluation of other potential measures of hydration status such as fingerstick P_{osm} (compared to venipuncture).

For the second problem, we develop a mathematical model to evaluate the impact of an electronic medical record system on the occurrence of medication errors and adverse drug events in a pediatric inpatient setting. In this chapter, we deal with count data that has a very large number of zeros and provide a methodology to model and select the most appropriate model. The first step in model selection is to obtain the mean-variance relationship of the data. If the variance is less than twice the mean, then the data is not overly dispersed and can be fit with a Poisson model. Next, if we have some knowledge about the presence of a group with "always" zeros, then the ZIP or HP should be evaluated. This methodology can be applied to other count data with excess zero, and the models can be extended by adding other covariates, including interaction terms.

This study also increases the knowledge base of implementing HITs by developing a better understanding of the impact of an incremental EMR implementation on patient safety, measured by the change in the rate of medical errors and adverse drug events. Our results of the data modeling indicate that HITs, in particular physician documentation and CPOE, reduce the rate of medication errors across hospitals and care areas. In the case of ADEs, the contribution of HIT is on the reduction of sources of variation of ADEs.

In the last chapter, we use predictive modeling to identify the risk factors and estimate a patient probability of developing post-operative nausea and vomiting based on

her demographics and clinical history. We develop several linear regression models to predict both, early and late PONV and then select the best one based on their AUC values and on their sensitivity-specificity combination. The latter criterion is particularly useful because it allows the evaluation of the models within specific ranges (e.g., be able to predict with at least 60% sensitivity). The predictors in our final models are not specific to the dataset used to develop the model, which is an advantage over other previously developed PONV models and scores. Another benefit of our models is that the significant predictors in both early and late PONV models are variables that can be easily gathered from the patient medical history before surgery to obtain the patient probability of developing PONV. These are: previous history of PONV, gender, ASA physical status, and obesity. For late PONV, early PONV was also a strong predictor. This is very relevant because if a patient did not have an episode of PONV during the first 2 hours, then it is most likely that she won't experience PONV in the next 22 hours. This factor had not been considered in any other PONV models.

REFERENCES

1. Armstrong, L.E., Hydration assessment techniques. Nutrition Reviews, 2005. 63(6): p. S40-S54.

2. Oppliger, R.A., et al., Accuracy of urine specific gravity and osmolality as indicators of hydration status. International Journal of Sport Nutrition and Exercise Metabolism, 2005. 15(3): p. 236-251.

3. Manz, F., Hydration and disease. Journal of the American College of Nutrition, 2007. 26(suppl 5): p. 535S-541S.

4. Manz, F. and A. Wentz, The importance of good hydration for the prevention of chronic diseases. Nutrition Reviews, 2005. 63: p. S2-S5.

5. Armstrong, L.E., Assessing hydration status: The elusive gold standard. J Am Coll Nutr, 2007. 26(Suppl 5): p. S575-S584.

6. Shirreffs, S.M., Markers of hydration status. European Journal of Clinical Nutrition, 2003. 57: p. S6-S9.

7. Minton, D.M. and L.E. Eberman, Best practice for clinical hydration measurement. Athletic Therapy Today, 2009. 14(1): p. 9-11.

8. Armstrong, L., et al., Urinary indices during dehydration, exercise, and rehydration. International Journal of Sport Nutrition, 1998. 8(4): p. 345-55.

9. Kavouras, S.A., Assessing hydration status. Current Opinion in Clinical Nutrition and Metabolic Care, 2002. 5(5): p. 519-524.

10. Popowski, L.A., et al., Blood and urinary measures of hydration status during progressive acute dehydration. Medicine and Science in Sports and Exercise, 2001. 33(5): p. 747-753.

11. Walsh, N.P., et al., Saliva parameters as potential indices of hydration status during acute dehydration. Medicine and Science in Sports and Exercise, 2004. 36(9): p. 1535-1542.

12. Yoshihara, A., et al., Serum markers of chronic dehydration are associated with saliva spinability. Journal of Oral Rehabilitation, 2007. 34(10): p. 733-738.

13. Oliver, S.J., et al., Saliva indices track hypohydration during 48 h of fluid restriction or combined fluid and energy restriction. Archives of Oral Biology, 2008. 53(10): p. 975-980.

14. Maughan, R.J., S.M. Shirreffs, and J.B. Leiper, Errors in the estimation of hydration status from changes in body mass. Journal of Sports Sciences, 2007. 25(7): p. 797-804.

15. King, R., et al., Estimating changes in hydration status from changes in body mass: Considerations regarding metabolic water and glycogen storage. Journal of Sports Sciences, 2008. 26(12): p. 1361-1363.

16. Baker, L., J. Lang, and W. Larry Kenney, Change in body mass accurately and reliably predicts change in body water after endurance exercise. European Journal of Applied Physiology, 2009. 105(6): p. 959-967.

17. Cheuvront, S.N., et al., Biological variation and diagnostic accuracy of dehydration assessment markers. The American Journal of Clinical Nutrition, 2010. 92(3): p. 565-573.

18. Hamouti, N., et al., Effects of athletes' muscle mass on urinary markers of hydration status. European Journal of Applied Physiology, 2010. 109(2): p. 213-219.

19. Maughan, R. and S. Shirreffs, Dehydration and rehydration in competative sport. Scandinavian journal of medicine & science in sports, 2010. 20: p. 40-47.

20. Wakefield, B., et al., Monitoring hydration status in elderly veterans. Western Journal of Nursing Research, 2002. 24(2): p. 132-142.

21. Armstrong, L.E., et al., Human hydration indices: Acute and longitudinal reference values. International Journal of Sport Nutrition and Exercise Metabolism, 2010. 20(2): p. 145-153.

22. Silva, R.P., et al., Preexercise urine specific gravity and fluid intake during one-hour running in a thermoneutral environment-a randomized cross-over study. Journal of Sports Science and Medicine, 2010. 9(3): p. 464-71.

23. Harvey, G., et al., The use of body mass changes as a practical measure of dehydration in team sports. Journal of Science and Medicine in Sport, 2008. 11(6): p. 600-603.

24. Miekisch, W. and J.K. Schubert, From highly sophisticated analytical techniques to life-saving diagnostics: Technical developments in breath analysis. Trac-Trends in Analytical Chemistry, 2006. 25(7): p. 665-673.

25. Cao, W.Q. and Y.X. Duan, Current status of methods and techniques for breath analysis. Critical Reviews in Analytical Chemistry, 2007. 37(1): p. 3-13.

26. Di Francesco, F., et al., Breath analysis: trends in techniques and clinical applications. Microchemical Journal, 2005. 79(1-2): p. 405-410.

27. Cao, W.Q. and Y.X. Duan, Breath analysis: Potential for clinical diagnosis and exposure assessment. Clinical Chemistry, 2006. 52(5): p. 800-811.

28. O'Hara, M.E., et al., Endogenous volatile organic compounds in breath and blood of healthy volunteers: examining breath analysis as a surrogate for blood measurements. Journal of Breath Research, 2009. 3(2).

29. Miekisch, W., J.K. Schubert, and G.F.E. Noeldge-Schomburg, Diagnostic potential of breath analysis - focus on volatile organic compounds. Clinica Chimica Acta, 2004. 347(1-2): p. 25-39.

30. Buszewski, B., et al., Human exhaled air analytics: Biomarkers of diseases. Biomedical Chromatography, 2007. 21(6): p. 553-566.

31. Probert, C.S.J., et al., Volatile organic compounds as diagnostic biomarkers in gastrointestinal and liver diseases. Journal of Gastrointestinal and Liver Diseases, 2009. 18(3): p. 337-343.

32. Szulejko, J.E., et al., Evidence for cancer biomarkers in exhaled breath. Ieee Sensors Journal, 2010. 10(1): p. 185-210.

33. Phillips, M., et al., Volatile organic compounds in breath as markers of lung cancer: a cross-sectional study. Lancet, 1999. 353(9168): p. 1930-1933.

34. Horvath, I., et al., Exhaled biomarkers in lung cancer. European Respiratory Journal, 2009. 34(1): p. 261-275.

35. Phillips, M., et al., Detection of lung cancer using weighted digital analysis of breath biomarkers. Clinica Chimica Acta, 2008. 393(2): p. 76-84.

36. Qin, T., et al., The screening of volatile markers for hepatocellular carcinoma. Cancer Epidemiology Biomarkers & Prevention, 2010. 19(9): p. 2247-2253.

37. Phillips, M., et al., Volatile biomarkers of pulmonary tuberculosis in the breath. Tuberculosis, 2007. 87(1): p. 44-52.

38. Dallinga, J., et al., Volatile organic compounds in exhaled breath as a diagnostic tool for asthma in children. Clinical & Experimental Allergy, 2010. 40(1): p. 68-76.

39. Fens, N., et al., Exhaled breath profiling enables discrimination of chronic obstructive pulmonary disease and asthma. American Journal of Respiratory and Critical Care Medicine, 2009. 180(11): p. 1076-1082.

40. Van Berkel, J., et al., A profile of volatile organic compounds in breath discriminates COPD patients from controls. Respiratory Medicine, 2010. 104(4): p. 557-563.

41. Greiter, M.B., et al., Differences in exhaled gas profiles between patients with type 2 diabetes and healthy controls. Diabetes Technology & Therapeutics, 2010. 12(6): p. 455-463.

42. Lee, J., et al., Improved predictive models for plasma glucose estimation from multi-linear regression analysis of exhaled volatile organic compounds. Journal of Applied Physiology, 2009. 107(1): p. 155-160.

43. Bennett, L., et al., A chemometric study on human breath mass spectra for biomarker identification in cystic fibrosis. Journal of Breath Research, 2009. 3(4).

44. Van Berkel, J., et al., Development of accurate classification method based on the analysis of volatile organic compounds from human exhaled air. Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences, 2008. 861(1): p. 101-107.

45. Ligor, T., Analytical methods for breath investigation. Critical Reviews in Analytical Chemistry, 2009. 39(1): p. 2-12.

46. O'Hara, M.E., et al., Mass spectrometric investigations to obtain the first direct comparisons of endogenous breath and blood volatile organic compound concentrations in healthy volunteers. International Journal of Mass Spectrometry, 2009. 281(1-2): p. 92-96.

47. King, J., et al., Dynamic profiles of volatile organic compounds in exhaled breath as determined by a coupled PTR-MS/GC-MS study. Physiological Measurement, 2010. 31(9): p. 1169-1184.

48. Armstrong, L.E., et al., Urinary indices of hydration status. Int J Sport Nutr, 1994. 4(3): p. 265-79.

49. Jolliffe, I.T. and MyiLibrary, Principal component analysis. Vol. 2. 2002: Wiley Online Library.

50. Dudoit, S., J. Fridlyand, and T.P. Speed, Comparison of discrimination methods for the classification of tumors using gene expression data. Journal of the American statistical association, 2002. 97(457): p. 77-87.

51. Wu, B., et al., Comparison of statistical methods for classification of ovarian cancer using mass spectrometry data. Bioinformatics, 2003. 19(13): p. 1636-1643.

52. Wagner, M., D. Naik, and A. Pothen, Protocols for disease classification from mass spectrometry data. Proteomics, 2003. 3(9): p. 1692-1698.

53. Datta, S. and V. Pihur, Feature selection and machine learning with mass spectrometry data. Methods in Molecular Biology, 2010. 593: p. 205-229.

54. Preventing Medication Errors, in Quality Chasm, Philip Aspden, et al., Editors. 2007, Institute of Medicine: Washington, DC. p. 463.

55. Barker, K.N., et al., Medication errors observed in 36 health care facilities. Arch Intern Med, 2002. 162(16): p. 1897-1903.

56. Bates, D., et al., Relationship between medication errors and adverse drug events. Journal of General Internal Medicine, 1995. 10(4): p. 199-205.

57. Kanjanarat, P., et al., Nature of preventable adverse drug events in hospitals: a literature review. American Journal of Health-System Pharmacy, 2003. 60(17): p. 1750-1759.

58. Devine, E.B., et al., Characterization of prescribing errors in an internal medicine clinic. American Journal Of Health-System Pharmacy: AJHP: Official Journal Of The American Society Of Health-System Pharmacists, 2007. 64(10): p. 1062-1070.

59. Evans, R.S., et al., Risk factors for adverse drug events: A 10-year analysis. Ann Pharmacother, 2005. 39(7): p. 1161-1168.

60. Morimoto, T., et al., Adverse drug events and medication errors: detection and classification methods. Quality and Safety in Health Care, 2004. 13(4): p. 306-314.

61. Wilmer, A., et al., Incidence of medication errors and adverse drug events in the ICU: a systematic review. Quality and Safety in Health Care, 2010. 19(5): p. 1.

62. Desikan, R., et al., Reporting of adverse drug events: examination of a hospital incident reporting system. Advance in patient safety: from research to implementation. 1: p. 1-4.

63. Kane-Gill, S.L., J.G. Kowiatek, and R.J. Weber, A comparison of voluntarily reported medication errors in intensive care and general care units. Quality and Safety in Health Care, 2010. 19(1): p. 55-59.

64. Hughes, R. and E. Edgerton, Reducing pediatric medication errors: children are especially at risk for medication errors. AJN The American Journal of Nursing, 2005. 105(5): p. 79-84.

65. Kaushal, R., K.N. Barker, and D.W. Bates, How can information technology improve patient safety and reduce medication errors in children's health care? Arch Pediatr Adolesc Med, 2001. 155(9): p. 1002-1007.

66. Kaushal, R., et al., Medication errors and adverse drug events in pediatric inpatients. JAMA, 2001. 285(16): p. 2114.

67. Agarwal, S., et al., Prevalence of adverse events in pediatric ICUs in the United States. Pediatric Critical Care Medicine, 2010. 11(5): p. 000-000.

68. Kopp, B.J., et al., Medication errors and adverse drug events in an intensive care unit: Direct observation approach for detection. Critical Care Medicines, 2006. 34(2): p. 415-425.

69. Holdsworth, M.T., et al., Incidence and impact of adverse drug events in pediatric inpatients. Arch Pediatr Adolesc Med, 2003. 157(1): p. 60-65.

70. DesRoches, C.M., et al., Small, nonteaching, and rural hospitals continue to be slow in adopting electronic health record systems. Health Affairs, 2012. 31(5): p. 1092-1099.

71. Ammenwerth, E., et al., The effect of electronic prescribing on medication errors and adverse drug events: a systematic review. Journal of the American Medical Informatics Association, 2008. 15(5): p. 585-600.

72. Jani, Y.H., et al., Electronic prescribing reduced prescribing errors in a pediatric renal outpatient clinic. The Journal of Pediatrics, 2008. 152(2): p. 214-218.

73. Bates, D., et al., The impact of computerized physician order entry on medication error prevention. Journal of the American Medical Informatics Association, 1999. 6(4): p. 313-321.

74. Agrawal, A., Medication errors: prevention using information technology systems. British Journal Of Clinical Pharmacology, 2009. 67(6): p. 681-686.

75. Bates, D.W., Using information technology to reduce rates of medication errors in hospitals. BMJ, 2000. 320(7237): p. 788-791.

76. Reckmann, M.H., et al., Does computerized provider order entry reduce prescribing errors for hospital inpatients? A systematic review. Journal of the American Medical Informatics Association, 2009. 16(5): p. 613-623.

77. Fortescue, E.B., et al., Prioritizing strategies for preventing medication errors and adverse drug events in pediatric inpatients. Pediatrics, 2003. 111(4): p. 722-729.

78. Devine, E.B., et al., The impact of computerized provider order entry on medication errors in a multispecialty group practice. Journal of the American Medical Informatics Association, 2010. 17(1): p. 78-84.

79. Holdsworth, M., et al., Impact of computerized prescriber order entry on the incidence of adverse drug events in pediatric inpatients. Pediatrics, 2007. 120(5): p. 1058-1066.

80. Hug, B., et al., Adverse drug event rates in six community hospitals and the potential impact of computerized physician order entry for prevention. Journal of General Internal Medicine, 2010. 25(1): p. 31-38.

81. Yu, F., et al., The relationship between computerized physician order entry and pediatric adverse drug events: a nested matched case-control study. Pharmacoepidemiology and Drug Safety, 2009. 18(8): p. 751-755.

82. Nebeker, J.R., et al., High rates of adverse drug events in a highly computerized hospital. Arch Intern Med, 2005. 165(10): p. 1111-1116.

83. Sard, B.E., et al., Retrospective evaluation of a computerized physician order entry adaptation to prevent prescribing errors in a pediatric emergency department. Pediatrics, 2008. 122(4): p. 782-787.

84. Shamliyan, T.A., et al., Just what the doctor ordered. review of the evidence of the impact of computerized physician order entry system on medication errors. Health Services Research, 2008. 43(1p1): p. 32-53.

85. Wang, J.K., et al., Prevention of pediatric medication errors by hospital pharmacists and the potential benefit of computerized physician order entry. Pediatrics, 2007. 119(1): p. e77-85.

86. Zlabek, J.A., J.W. Wickus, and M.A. Mathiason, Early cost and safety benefits of an inpatient electronic health record. Journal of the American Medical Informatics Association, 2011. 18(2): p. 169-172.

87. Forni, A., H.T. Chu, and J. Fanikos, Technology utilization to prevent medication errors. Current Drug Safety, 2010. 5(1): p. 13-18.

88. Mekhjian, H., et al., Immediate benefits realized following implementation of physician order entry at an academic medical center. Journal of the American Medical Informatics Association, 2002. 9(5): p. 529.

89. Hassan, E., et al., Using technology to prevent adverse drug events in the intensive care unit. Critical Care Medicine, 2010. 38: p. S97-S105 10.1097/CCM.0b013e3181dde1b4.

90. van Rosse, F., et al., The effect of computerized physician order entry on medication prescription errors and clinical outcome in pediatric and intensive care: A systematic review. Pediatrics, 2009. 123(4): p. 1184-1190.

91. Buntin, M.B., et al., The benefits of health information technology: a review of the recent literature shows predominantly positive results. Health Affairs, 2011. 30(3): p. 464-471.

92. Koppel, R., et al., Role of computerized physician order entry systems in facilitating medication errors. JAMA, 2005. 293(10): p. 1197-1203.

93. Zhan, C., et al., Potential benefits and problems with computerized prescriber order entry: analysis of a voluntary medication error-reporting database. American Journal of Health-System Pharmacy, 2006. 63(4): p. 353-358.

94. Piontek, F., et al., Effects of an adverse-drug-event alert system on cost and quality outcomes in community hospitals. American Journal Of Health-System Pharmacy, 2010. 67(8): p. 613-620.

95. Zandieh, S.O., et al., Risk factors in preventable adverse drug events in pediatric outpatients. The Journal of Pediatrics, 2008. 152(2): p. 225-231.

96. Tissot, E., et al., Observational study of potential risk factors of medication administration errors. Pharmacy World & Science, 2003. 25(6): p. 264-268.

97. Bates, D.W., et al., Patient risk factors for adverse drug events in hospitalized patients. Arch Intern Med, 1999. 159(21): p. 2553-2560.

98. Wilkins, K. and M. Shields, Correlates of medication error in hospitals. Health Reports, 2008. 19(2): p. 7-18.

99. Gagnon, D., et al., Poisson regression for modeling count and frequency outcomes in trauma research. Journal of Traumatic Stress, 2008. 21(5): p. 448-454.

100. Byers, A.L., et al., Application of negative binomial modeling for discrete outcomes:: A case study in aging research. Journal of clinical epidemiology, 2003. 56(6): p. 559-564.

101. Hur, K., et al., Modeling clustered count data with excess zeros in health care outcomes research. Health Services and Outcomes Research Methodology, 2002. 3(1): p. 5-20.

102. Cheung, Y., Zero-inflated models for regression analysis of count data: a study of growth and development. Statistics in medicine, 2002. 21(10): p. 1461-1469.

103. Cameron, A.C. and P.K. Trivedi, Regression analysis of count data. 1998: Cambridge University Press.

104. Takata, G.S., C.K. Taketomo, and S. Waite, Characteristics of medication errors and adverse drug events in hospitals participating in the California Pediatric Patient Safety Initiative. American Journal of Health-System Pharmacy, 2008. 65(21): p. 2036-2044.

105. Takata, G.S., et al., Development, Testing, and Findings of a Pediatric-Focused Trigger Tool to Identify Medication-Related Harm in US Children's Hospitals. Pediatrics, 2008. 121(4): p. e927-935.

106. Rozich, J.D., C.R. Haraden, and R.K. Resar, Adverse drug event trigger tool: a practical methodology for measuring medication related harm. Quality and Safety in Health Care, 2003. 12(3): p. 194-200.

107. Bobb, A., et al., The epidemiology of prescribing errors: The potential impact of computerized prescriber order entry. Arch Intern Med, 2004. 164(7): p. 785-792.

108. Collins, C.M. and K.A. Elsaid, Using an enhanced oral chemotherapy computerized provider order entry system to reduce prescribing errors and

improve safety. International Journal for Quality in Health Care, 2011. 23(1): p. 36-43.

109. Cordero, L., et al., Impact of computerized physician order entry on clinical practice in a newborn intensive care unit. Journal of Perinatology, 2004. 24(2): p. 88-93.

110. Mahoney, C.D., et al., Effects of an integrated clinical information system on medication safety in a multi-hospital setting. American Journal of Health-System Pharmacy, 2007. 64(18): p. 1969-1977.

111. Warton, D., Many zeros does not mean zero inflation: comparing the goodness-of-fit of parametric models to multivariate abundance data. Environmetrics, 2005. 16(3): p. 275-289.

112. Neuhauser, D., L. Provost, and B. Bergman, The meaning of variation to healthcare managers, clinical and health-services researchers, and individual patients. BMJ Quality & Safety, 2011. 20(Suppl 1): p. i36-i40.

113. Habib, A.S., et al., Postoperative nausea and vomiting following inpatient surgeries in a teaching hospital: a retrospective database analysis. Current Medical Research and Opinion, 2006. 22(6): p. 1093-1099.

114. Kreis, M., Postoperative nausea and vomiting. Autonomic Neuroscience, 2006. 129(1-2): p. 86-91.

115. Williams, K.S., Postoperative nausea and vomiting. Surgical Clinics of North America, 2005. 85(6): p. 1229-41.

116. Miller, T., A. Roche, and T. Gan, Definitions and epidemiology. Current Opinion in Critical Care, 2009. 15(4): p. 314-319.

117. Kovac, A.L., Prevention and Treatment of Postoperative Nausea and Vomiting. Drugs, 2000. 59(2): p. 213-243.

118. Borgeat, A., G. Ekatodramis, and C.A. Schenker, Postoperative nausea and vomiting in regional anesthesia - A review. Anesthesiology, 2003. 98(2): p. 530-547.

119. Brattwall, M., et al., Postoperative impact of regular tobacco use, smoking or snuffing, a prospective multi-center study. Acta Anaesthesiologica Scandinavica, 2010. 54(3): p. 321-327.

120. Kranke, P., et al., An increased body mass index is no risk factor for postoperative nausea and vomiting - A systematic review and results of original data. Acta Anaesthesiologica Scandinavica, 2001. 45(2): p. 160-166.

121. Gan, T.J., Risk factors for postoperative nausea and vomiting. Anesthesia & Analgesia, 2006. 102(6): p. 1884-1898.

122. Habib, A.S. and T.J. Gan, CON: Postoperative nausea and vomiting database research: Limitations and opportunities. Anesthesia and Analgesia, 2010. 110(2): p. 412-414.

123. Apfel, C., et al., Volatile anaesthetics may be the main cause of early but not delayed postoperative vomiting: a randomized controlled trial of factorial design. British Journal of Anaesthesia, 2002. 88(5): p. 659.

124. Nakagawa, M., et al., Dopamine D2 receptor Taq IA polymorphism is associated with postoperative nausea and vomiting. Journal of Anesthesia, 2008. 22(4): p. 397-403.

125. Nitahara, K., et al., Risk factors for nausea and vomiting following vitrectomy in adults. European Journal of Anaesthesiology, 2007. 24(2): p. 166-170.

126. Choi, D.H., et al., A Korean predictive model for postoperative nausea and vomiting. Journal of Korean Medical Science, 2005. 20(5): p. 811-815.

127. Stadler, M., et al., Difference in risk factors for postoperative nausea and vomiting. Anesthesiology, 2003. 98(1): p. 46-52.

128. Kranke, P., F. Schuster, and L.H. Eberhart, Recent advances, trends and economic considerations in the risk assessment, prevention and treatment of postoperative nausea and vomiting. Expert Opinion on Pharmacotherapy, 2007. 8(18): p. 3217-3235.

129. Rüsch, D., et al., Prospective application of a simplified risk score to prevent postoperative nausea and vomiting. Canadian Journal of Anesthesia/Journal canadien d'anesthésie, 2005. 52(5): p. 478-484.

130. Golembiewski, J., E. Chernin, and T. Chopra, Prevention and treatment of postoperative nausea and vomiting. American Journal of Health-System Pharmacy, 2005. 62(12): p. 1247-62.

131. Biedler, A., et al., A risk adapted approach reduces the overall institutional incidence of postoperative nausea and vomiting. Canadian Journal of Anaesthesia-Journal Canadien D Anesthesie, 2004. 51(1): p. 13-19.

132. Kloth, D.D., New pharmacologic findings for the treatment of PONV and PDNV. American Journal of Health-System Pharmacy, 2009. 66: p. S11-S18.

133. Percival, V.G., J. Riddell, and T.B. Corcoran, Single dose dexamethasone for postoperative nausea and vomiting - a matched case-control study of postoperative infection risk. Anaesthesia and Intensive Care, 2010. 38(4): p. 661-666.

134. Ignoffo, R.J., Current research on PONV/PDNV: Practical implications for today's pharmacist. American Journal of Health-System Pharmacy, 2009. 66(1): p. S19-S24.

135. Apfel, C., et al., A factorial trial of six interventions for the prevention of postoperative nausea and vomiting. New England Journal of Medicine, 2004. 350(24): p. 2441-2451.

136. Gan, T.J., et al., Society for ambulatory anesthesia guidelines for the management of postoperative nausea and vomiting. Anesthesia and Analgesia, 2007. 105(6): p. 1615-1628.

137. Rüsch, D., et al., Nausea and vomiting after surgery under general anesthesia an evidence-based review concerning risk assessment, prevention, and treatment. Deutsches Arzteblatt International, 2010. 107(42): p. 733-41.

138. Habib, A. and T. Gan, Evidence-based management of postoperative nausea and vomiting: a review. Canadian Journal of Anesthesia/Journal canadien d'anesthésie, 2004. 51(4): p. 326-341.

139. Wilhelm, S.M., M.L. Dehoorne-Smith, and P.B. Kale-Pradhan, Prevention of postoperative nausea and vomiting. The Annals of Pharmacotherapy, 2007. 41(1): p. 68-78.

140. Lee, Y.Y., K.H. Kim, and Y.H. Yom, Predictive models for post-operative nausea and vomiting in patients using patient-controlled analgesia. Journal of International Medical Research, 2007. 35(4): p. 497-507.

141. Peng, S.Y., et al., Predicting postoperative nausea and vomiting with the application of an artificial neural network. British Journal of Anaesthesia, 2007. 98(1): p. 60-65.

142. Apfel, C., et al., A simplified risk score for predicting postoperative nausea and vomiting: conclusions from cross-validations between two centers. Anesthesiology, 1999. 91(3): p. 693.

143. White, P.F., et al., The relationship between patient risk factors and early versus late postoperative emetic symptoms. Anesthesia and Analgesia, 2008. 107(2): p. 459-463.

144. Koivuranta, M., et al., A survey of postoperative nausea and vomiting. Anaesthesia, 1997. 52(5): p. 443-449.

145. Van den Bosch, J., et al., Assessing the applicability of scoring systems for predicting postoperative nausea and vomiting. Anaesthesia, 2005. 60(4): p. 323-331.

146. Thomas, R., N.A. Jones, and P. Strike, The value of risk scores for predicting postoperative nausea and vomiting when used to compare patient groups in a randomised controlled trial. Anaesthesia, 2002. 57(11): p. 1119-1128.

147. Kapoor, R., et al., Comparison of two instruments for assessing risk of postoperative nausea and vomiting. American Journal of Health-System Pharmacy, 2008. 65(5): p. 448-453.

148. Pierre, S., H. Benais, and J. Pouymayou, Apfel's simplified score may favourably predict the risk of postoperative nausea and vomiting. Canadian Journal of Anaesthesia-Journal Canadien D Anesthesie, 2002. 49(3): p. 237-242.

149. Eberhart, L., et al., Evaluation of three risk scores to predict postoperative nausea and vomiting. Acta Anaesthesiologica Scandinavica, 2000. 44(4): p. 480-488.

150. Apfel, C.C., et al., Comparison of predictive models for postoperative nausea and vomiting. British Journal of Anaesthesia, 2002. 88(2): p. 234-240.

151. Apfel, C.C., et al., What can be expected from risk scores for predicting postoperative nausea and vomiting? British Journal of Anaesthesia, 2001. 86(6): p. 822-827.

152. Palazzo, M. and R. Evans, Logistic regression analysis of fixed patient factors for postoperative sickness: a model for risk assessment. British Journal of Anaesthesia, 1993. 70(2): p. 135-140.

153. Toner, C.C., et al., Prediction of postoperative nausea and vomiting using a logistic regression model. British Journal of Anaesthesia, 1996. 76(3): p. 347-351.

154. Sinclair, D.R., F. Chung, and G. Mezei, Can postoperative nausea and vomiting be predicted? Anesthesiology, 1999. 91(1): p. 109-118.

155. Velickovic, J., MD, Possibilities for prevention of postoperative nausea and vomiting in thyroid surgery, in Medical School. 2009, University of Belgrade, Serbia: Belgrade.

156. Glass, P.S.A., Postoperative nausea and vomiting: we don't know everything yet. Anesthesia and Analgesia, 2010. 110(2): p. 299-299.

157. Apfel, C.C. and E. George, Myths hardly ever die, but even if they do, one cannot be sure about it. Journal of Clinical Anesthesia, 2010. 22(7): p. 489-491.

VITA

SOFIA E. ESPINOZA

Sofia Elizabeth Espinoza was born in Lima, Peru. She earned her B.S. in Electrical Engineering from the Pontifical Catholic University of Peru in 1998 and her M.S. in Health Systems from Georgia Tech in 2004. She continued at Georgia Tech to pursue a doctorate in Industrial & Systems Engineering (ISyE). Her research interests focus on the application of statistical methods in healthcare research, and include data mining and classification methods, predictive modeling and evaluation of the impact of healthcare information technologies (HIT) on healthcare outcomes. She has worked in projects with Children's Healthcare of Atlanta (CHOA) and the Aflac Cancer Center, where she was involved in the evaluation of an HIT application to support ambulatory pediatric cancer survivor care.